



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 116176

TO: Phillip Gambel
Location: 3e81/3c70
Tuesday, March 09, 2004
Art Unit: 1644
Phone: 272-0844
Serial Number: 10 / 042421

From: Jan Delaval
Location: Biotech-Chem Library
Rem 1A51
Phone: 272-2504
jan.delaval@uspto.gov

Search Notes

10/042421

SEARCH REQUEST FORM

116176

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 3/9/04
Searcher: Jan
Terminal time: _____
Elapsed time: 12 + 10
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site

☒ STIC
☐ CM-1
☐ Pre-S

Type of Search

☐ N.A. Sequence
☒ A.A. Sequence
☐ Structure
☐ Bibliographic

Vendors

☐ IG
☐ STN
☐ Dialog
☐ APS
☐ Geninfo
☐ SDC
☒ DARC/Questel
☐ Other

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 21 Seconds
(without alignment)

2258.211 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 2608
Sequence: 1 MDKFWHAAWGLCLVPLSLA.....QFMADETRLQNVDMKIGV 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2590	99.3	493	2	CD44E protein, epi
2	2490	95.5	699	2	epican - human
3	2468.5	94.7	742	2	lymphocyte homing
4	2261	86.7	426	2	lymphocyte homing
5	2036	78.1	395	2	CD44R5 - human
6	1835	70.4	351	2	cell adhesion mole
7	1698.5	65.1	352	2	lymphocyte adhesio
8	1547	59.3	359	2	lymphocyte surface
9	1470	56.4	351	2	CD44 antigen precu
10	1469.5	56.3	366	2	cell-surface glyco
11	1427.5	54.7	362	2	T-cell surface gly
12	1425	54.6	503	2	cell adhesion mole
13	1368	52.5	363	2	CD44 homolog membr
14	1361	52.2	365	2	CD44 membrane gly
15	1361	52.2	365	2	CD44 membrane gly
16	699	26.8	337	2	CD44 protein - hum
17	541.5	20.8	146	2	cell surface glyco
18	525	20.1	148	2	cell surface glyco
19	394	15.1	301	2	CD44 glycoprotein
20	382.5	14.7	142	2	CD44 glycoprotein
21	380	14.6	414	2	CD44 protein - mou
22	344	13.2	70	2	cell surface glyco
23	233.5	9.0	107	2	CD44 glycoprotein
24	183	7.0	2232	2	hypothetical prote
25	178	6.8	833	2	lipoprotein vas
26	171	6.6	276	2	TSG-6 homolog
27	168	6.4	2124	2	proteoglycan core
28	163	6.3	786	2	hypothetical prote
29	162.5	6.2	275	2	tumor necrosis fac
			277	2	hyaluronate-bindin

30	161.5	6.2	631	2	epistatin - mouse
31	161	6.2	1275	2	hypothetical prote
32	158	6.1	630	2	tumor-associated m
33	156.5	6.0	1459	2	hypothetical prote
34	156	6.0	1032	2	hypothetical prote
35	155.5	6.0	15182	1	aggreacan precursor
36	152.5	5.8	1643	2	aggreacan precursor
37	152.5	5.8	1643	2	aggreacan precursor
38	152.5	5.8	3381	2	aggreacan precursor
39	150.5	5.8	2109	1	aggreacan precursor
40	150	5.8	636	2	probable membrane
41	150	5.8	862	2	secreted acid phos
42	149	5.7	1217	2	secreted acid phos
43	148.5	5.7	2271	2	hypothetical prote
44	148.5	5.7	2397	1	versican precursor
45	148	5.7	1367	1	glucan 1,4-alpha-g

ALIGNMENTS

Result 1	CD44E protein, epithelial - human
C/Species: Homo sapiens (man)	
C/Date: 18-Feb-1994	#sequence _revision 10-Nov-1995 #text_change 21-Jul-2000
C/Accession: S13530	
R/Stamenkovic, I.; Aruffo, A.; Amiot, M.; Seed, B.	
A/Title: The hematopoietic and epithelial forms of CD44 are distinct polypeptides w	
A/Reference number: S13530; MUID:91122041; PMID:1991450	
A/Accession: S13530	
A/Status: preliminary	
A/Molecule type: mRNA	
A/Residues: 1-493 <ST>	
A/Cross-references: EMBL:X5150; NID:929800; PIDN:CAA38951.1; PID:929801	
C/Keywords: transmembrane protein	
Query Match	99.3%; Score 2590; DB 2; Length 493;
Best Local Similarity	99.4%; Pred. No. 4,7e-159;
Matches 490; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	1 MDKFWHAAWGLCLVPLSLAQIDINTCRAGVHVKNGRYSISREADLCAPNSTL 60
DB	1 MDKFWHAAWGLCLVPLSLAQIDINTCRAGVHVKNGRYSISREADLCAPNSTL 60
QY	61 PTMAQMEKALSIGFETCRVYGBGVVPIPIHNSICANNVTGYIITSTGYDYCFN 120
DB	61 PTMAQMEKALSIGFETCRVYGBGVVPIPIHNSICANNVTGYIITSTGYDYCFN 120
QY	121 ASAPPEEDCTVMDLPNAPGPTITTVNDGRRVYKGRVTNPEIYSPNPDVVSS 180
DB	121 ASAPPEEDCTVMDLPNAPGPTITTVNDGRRVYKGRVTNPEIYSPNPDVVSS 180
QY	121 ASAPPEEDCTVMDLPNAPGPTITTVNDGRRVYKGRVTNPEIYSPNPDVVSS 180
DB	121 ASAPPEEDCTVMDLPNAPGPTITTVNDGRRVYKGRVTNPEIYSPNPDVVSS 180
QY	181 GSSERSSTSGCYIFFTFTVPIPDSDPWITDTRIPATNDSHSTTLOPTANPT 240
DB	181 GSSERSSTSGCYIFFTFTVPIPDSDPWITDTRIPATNDSHSTTLOPTANPT 240
QY	241 GLVEDLDRTGPLSTTQOSNSQSFSTHGLEBKDPHTSTLSSNRNVTGGRPNH 300
DB	241 GLVEDLDRTGPLSTTQOSNSQSFSTHGLEBKDPHTSTLSSNRNVTGGRPNH 300
QY	301 SEGSTHLLLEGYTSHPHTKSRTEIPYTSKATGSGFYTAATVGDNSNVRSLSGDDTF 360
DB	301 SEGSTHLLLEGYTSHPHTKSRTEIPYTSKATGSGFYTAATVGDNSNVRSLSGDDTF 360
QY	361 HPSGCSRTTHSSEDSHSGSQEGGANTTGPRTPIPEMLIILASLLALILANCA 420
DB	361 HPSGCSRTTHSSEDSHSGSQEGGANTTGPRTPIPEMLIILASLLALILANCA 420
QY	421 VNSRRRCGQKKLVINGNGAVEDRKSGSLNGEASKSQEWVHLVKNKSSSTPDQFWTAD 480
DB	421 VNSRRRCGQKKLVINGNGAVEDRKSGSLNGEASKSQEWVHLVKNKSSSTPDQFWTAD 480

QY 481 TRNLQNDMKIGV 493
Db 481 TRNLQNDMKIGV 493

RESULT 2

137369
epican - human
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Aug-2001
C:Accession: 137369; S24631
R:Kugelman, L.C.; Ganguly, S.; Haggerty, J.G.; Weissman, S.M.; Milstone, L.M.
J. Invest. Dermatol. 99, 866-891, 1992
A:Title: The core protein of epican, a heparan sulfate proteoglycan on keratinocytes, is
A:Reference number: 137369; PMID:1281868
A:Note: corrected and republished from J. Invest. Dermatol. 99, 381-385, 1992
A:Accession: 137369
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-699 <RES>
A:Cross-references: EMBL:X66733; NID:g31190; PID:g31191

Query Match 95.5%; Score 2490; DB 2; Length 699;
Best Local Similarity 70.4%; Pred. No. 2e-152;
Matches 492; Conservative 1; Mismatches 0; Indels 206; Gaps 1;

QY 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
Db 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
QY 61 PTMAQMERKALSIGFETCRGFFIEGHVVPRIHPNSICANNNGVYILTSNTSOYDYTCFN 120
Db 61 PTMAQMERKALSIGFETCRGFFIEGHVVPRIHPNSICANNNGVYILTSNTSOYDYTCFN 120
QY 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
Db 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
QY 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
Db 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
QY 181 GSSSRSSSTSGGYIYFTSTYVHPIDEDSPWITDSTRIPAT----- 222
Db 181 GSSSRSSSTSGGYIYFTSTYVHPIDEDSPWITDSTRIPAT----- 222
QY 223 ----- 222
Db 223 ----- 222
QY 241 DERDRHLSFGSGIDDDDFISSTISTTPRAFDHTKQNDWTQNNPSHNPVLLQTTTR 300
Db 241 DERDRHLSFGSGIDDDDFISSTISTTPRAFDHTKQNDWTQNNPSHNPVLLQTTTR 300
QY 223 ----- 222
Db 223 ----- 222
QY 301 MTVDVNGTAYEGNWNPEAPPLIHHEHEEETPHSTSTIOATPSSYTEETATQXQW 360
Db 301 MTVDVNGTAYEGNWNPEAPPLIHHEHEEETPHSTSTIOATPSSYTEETATQXQW 360
QY 223 ----- 222
Db 223 ----- 222
QY 361 FGNRWAVGRQTPKXDSSTGTGAASAHTSHPMQGRTPSPEDSSWTFNPIISHMGR 420
Db 361 FGNRWAVGRQTPKXDSSTGTGAASAHTSHPMQGRTPSPEDSSWTFNPIISHMGR 420
QY 223 ----- 222
Db 223 ----- 222
QY 421 GHQAGRMDMDSSHTTLQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEED 480
Db 421 GHQAGRMDMDSSHTTLQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEED 480
QY 275 KQHPTSTLTSNRNDVYGGRRDPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGS 334
Db 275 KQHPTSTLTSNRNDVYGGRRDPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGS 334
QY 481 KQHPTSTLTSNRNDVYGGRRDPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGS 540
Db 481 KQHPTSTLTSNRNDVYGGRRDPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGS 540
QY 335 FGVTAVTVGDSNSNVRSLSGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIR 394
Db 335 FGVTAVTVGDSNSNVRSLSGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIR 394
QY 541 FGVTAVTVGDSNSNVRSLSGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIR 600
Db 541 FGVTAVTVGDSNSNVRSLSGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIR 600
QY 395 TQOIPBWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 454
Db 395 TQOIPBWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 454
QY 601 TQOIPBWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 660
Db 601 TQOIPBWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 660
QY 455 SKSGEMVHLVKNKESSETPDQFMFADETRNQLQNDMKIGV 493
Db 455 SKSGEMVHLVKNKESSETPDQFMFADETRNQLQNDMKIGV 493

Db 661 SKSGEMVHLVKNKESSETPDQFMFADETRNQLQNDMKIGV 699

RESULT 3

A47195
lymphocyte homing receptor isoform CD44 - human
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: A47195
R:Scranton, G.R.; Bell, M.V.; Jackson, D.G.; Cornelius, F.B.; Gerth, U.; Bell, J.I.
Proc. Natl. Acad. Sci. U.S.A. 89, 12160-12164, 1992
A:Title: Genomic structure of DNA encoding the lymphocyte homing receptor CD44 revea
A:Reference number: A47195; MUID:93101687; PMID:1465456
A:Accession: A47195
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-742 <SCR>
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI Dackbone (NCBIN:120731, NCBIN:120737, NCBIN:12
, NCBIN:120764, NCBIN:120766, NCBIN:120770, NCBIN:120772, NCBIN:120774, NCBIN:120777,

Query Match 94.7%; Score 2468.5; DB 2; Length 742;
Best Local Similarity 66.3%; Pred. No. 5.1e-151;
Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

QY 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
Db 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
QY 61 PTMAQMERKALSIGFETCRGFFIEGHVVPRIHPNSICANNNGVYILTSNTSOYDYTCFN 120
Db 61 PTMAQMERKALSIGFETCRGFFIEGHVVPRIHPNSICANNNGVYILTSNTSOYDYTCFN 120
QY 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
Db 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
QY 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
Db 121 ASAPPEEDCTSTYDLPNAPFDGPIITTYVNRDGTFRVQGEKRTNPEDIYPSNPDDVSS 180
QY 181 GSSSRSSSTSGGYIYFTSTYVHPIDEDSPWITDSTRIPAT----- 222
Db 181 GSSSRSSSTSGGYIYFTSTYVHPIDEDSPWITDSTRIPAT----- 222
QY 223 ----- 222
Db 223 ----- 222
QY 241 TWDPSWFLPESRNHLHTTSSNTISAGWEPNEEDTQWAGERDRLSFGSGIDDD 300
Db 241 TWDPSWFLPESRNHLHTTSSNTISAGWEPNEEDTQWAGERDRLSFGSGIDDD 300
QY 223 ----- 222
Db 223 ----- 222
QY 301 EDPISSTITPRADHTKQNDWTQNNPSHNPVLLQTTTRMTVDVNRNGTAYEGNWN 360
Db 301 EDPISSTITPRADHTKQNDWTQNNPSHNPVLLQTTTRMTVDVNRNGTAYEGNWN 360
QY 223 ----- 222
Db 223 ----- 222
QY 361 PEAPPLIHHEHEEETPHSTSTIOATPSSYTEETATQXQWFGNWHGGRQTPRDS 420
Db 361 PEAPPLIHHEHEEETPHSTSTIOATPSSYTEETATQXQWFGNWHGGRQTPRDS 420
QY 223 ----- 222
Db 223 ----- 222
QY 421 HSTTGAASAHTSHPMQGRTPSPEDSSWTFNPIISHMGRGHQAGRMDMDSSHT 480
Db 421 HSTTGAASAHTSHPMQGRTPSPEDSSWTFNPIISHMGRGHQAGRMDMDSSHT 480
QY 232 LQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 291
Db 232 LQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 291
QY 481 LQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 540
Db 481 LQPTANPMTGLVEDIDRTGPLSMTOOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 540
QY 292 TGRDRPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGSFGVTAVTVGDSNSNVR 351
Db 292 TGRDRPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGSFGVTAVTVGDSNSNVR 351
QY 541 TGRDRPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGSFGVTAVTVGDSNSNVR 600
Db 541 TGRDRPNHSRSGSTTLLEGYTSYHPHTKESRFTIPIVTSKATGSFGVTAVTVGDSNSNVR 600
QY 352 SLGGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIRTPQIPBWLIIILASLAL 411
Db 352 SLGGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIRTPQIPBWLIIILASLAL 411
QY 601 SLGGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIRTPQIPBWLIIILASLAL 660
Db 601 SLGGDDTFHPSGGSHHTTGSBDGSHSGOEGGANTTSGPIRTPQIPBWLIIILASLAL 660
QY 412 ALLAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEASKSGEMVHLVKNKESSE 471
Db 412 ALLAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEASKSGEMVHLVKNKESSE 471
QY 661 ALLAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEASKSGEMVHLVKNKESSE 720
Db 661 ALLAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEASKSGEMVHLVKNKESSE 720

QY 472 PDQFMTADERTNLQNVDMKIGV 493
 |||||
 Db 721 PDQFMTADERTNLQNVDMKIGV 742

RESULT 4

JH0518
 Lymphocyte homing receptor CD44, splice form CD44R1 - human
 N:Alternate names: cell adhesion molecule core protein CD44E, keratinocyte, cell surface
 N:Contains: Lymphocyte homing receptor CD44, splice form CD44R1, Lymphocyte homing recep
 C:Species: Homo sapiens (man)
 C>Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text change 18-Aug-2000
 C:Accession: JH0518; JH0519; PH0859; A39209; A42402; C42402; A53029; S16147
 R:Dougherty, G.J.; Landsorp, P.M.; Cooper, D.L.; Humphries, R.K.
 J. Exp. Med. 174, 1-5, 1991
 A:Title: Molecular cloning of CD44R1 and CD44R2, two novel isoforms of the human CD44 ly
 A:Reference number: JH0518; MUID:91277598; PMID:2056274
 A:Accession: JH0518
 A:Molecule type: mRNA
 A:Residues: 1-426 <DOU>
 A:Experimental source: Lymphocytes, cell line KG1a
 A:Accession: JH0519
 A:Molecule type: mRNA
 A:Residues: 1-223,288-426 <DO2>
 A:Experimental source: Lymphocyte, cell line KG1a
 R:Cooper, D.L.; Dougherty, G.; Harn, H.J.; Jackson, S.; Baptist, E.W.; Byers, J.; Datta,
 Biochem. Biophys. Res. Commun. 182, 569-578, 1992
 A:Title: The complex CD44 transcritpional unit: alternative splicing of three internal e
 A:Reference number: PH0859; MUID:92134271; PMID:1734871
 A:Accession: PH0859
 A:Molecule type: DNA
 A:Residues: 223-357 <COO>
 R:Brown, T.A.; Bouchard, T.; St. John, T.; Wayner, E.; Carter, W.G.
 J. Cell Biol. 113, 207-221, 1991
 A:Title: Human Keratinocytes express a new CD44 core protein (CD44E) as a heparan-sulfat
 A:Reference number: A39209; MUID:91177958; PMID:2007624
 A:Accession: A39209
 A:Molecule type: mRNA
 A:Residues: 184-376 <BRO>
 A:Cross-references: GB:X55938; NID:929802; PIDN:CAA39404.1; PID:93930047
 R:Jackson, D.G.; Buckley, J.F.; Bell, J.T.
 J. Biol. Chem. 267, 4732-4739, 1992
 A:Title: Multiple variants of the human lymphocyte homing receptor CD44 generated by ins
 A:Reference number: A42402; MUID:92165834; PMID:1537855
 A:Accession: A42402
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 217-223,288-359 <UNC>
 A:Note: sequence extracted from NCBI backbone (NCBI:N:83964, NCBI:P:83965)
 A:Accession: C42402
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 217-370 <JAZ>
 A:Note: sequence extracted from NCBI backbone (NCBI:N:83968, NCBI:P:83969)
 A:Accession: A53029
 A:Status: preliminary
 A:Molecule type: Protein
 A:Residues: 67-76, 'X', 78-89 <SHB>
 C:Genetics:
 A:Gene: GDB:CD44; MDU2; MDU3; MI
 A:Cross-references: GDB:120739; OMIM:107269
 A:Map position: 11pter-11p13
 A:Introns: 35/1; 65/1; 133/1
 C:Superfamily: human cell adhesion protein CD44
 C:Keywords: alternative splicing; cell adhesion; chondroitin sulfate proteoglycan; glyco
 F:1-426/Product: Lymphocyte homing receptor CD44, splice form CD44R1 #status predicted <
 F:1-223,288-426/Product: Lymphocyte homing receptor CD44, splice form CD44R2 #status pre

F:299/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:354/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.7%; Score 2261; DB 2; Length 426;
 Best Local Similarity 100.0%; Pred. No. 5,4e-138;
 Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKFWMAWAGLCVPLSLAQIDINTCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60
 |||||
 Db 1 MDKFWMAWAGLCVPLSLAQIDINTCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60
 QY 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120
 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120
 Db 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120
 QY 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGRVYQKGEYRNPEDIYSPNPTDDVSS 180
 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGRVYQKGEYRNPEDIYSPNPTDDVSS 180
 Db 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGRVYQKGEYRNPEDIYSPNPTDDVSS 180
 QY 181 GSSSRSSSTSGYIFFTSTYHPIDEDSPMITTSTDRIPATNMDSHSTLLOFTANPT 240
 181 GSSSRSSSTSGYIFFTSTYHPIDEDSPMITTSTDRIPATNMDSHSTLLOFTANPT 240
 Db 181 GSSSRSSSTSGYIFFTSTYHPIDEDSPMITTSTDRIPATNMDSHSTLLOFTANPT 240
 QY 241 GLVEDLDRTPGLSMVTQSNQSFSTHEGLEEDKDPHTTSTLTSNRNDVYQGRDPNH 300
 241 GLVEDLDRTPGLSMVTQSNQSFSTHEGLEEDKDPHTTSTLTSNRNDVYQGRDPNH 300
 Db 241 GLVEDLDRTPGLSMVTQSNQSFSTHEGLEEDKDPHTTSTLTSNRNDVYQGRDPNH 300
 QY 301 SBGSTLLLEGYTSHPHTKESRTPIPTVSATGSPGYTAVYGSNSNVNLSLGDQDTF 360
 301 SBGSTLLLEGYTSHPHTKESRTPIPTVSATGSPGYTAVYGSNSNVNLSLGDQDTF 360
 Db 301 SBGSTLLLEGYTSHPHTKESRTPIPTVSATGSPGYTAVYGSNSNVNLSLGDQDTF 360
 QY 361 HPSGGSHTTHSGESDGHSGQEGCAVNTSPITPQIPFWLIIILALALILAVCIA 420
 361 HPSGGSHTTHSGESDGHSGQEGCAVNTSPITPQIPFWLIIILALALILAVCIA 420
 Db 361 HPSGGSHTTHSGESDGHSGQEGCAVNTSPITPQIPFWLIIILALALILAVCIA 420
 QY 421 VNSRR 425
 421 VNSRR 425
 Db 421 VNSRR 425

RESULT 5
 177371
 CD44R5 - human
 C:Species: Homo sapiens (man)
 C>Date: 02-Aug-1996 #sequence revision 02-Aug-1996 #text change 21-Jul-2000
 C:Accession: I77371
 R:Farabee, K.K.; Nishi, T.; Saya, H.
 Mol. Carcinog. 7, 212-220, 1993
 A:Title: Novel variants of CD44 arising from alternative splicing: changes in the CI
 A:Reference number: I57483; MUID:93356912; PMID:8352881
 A:Accession: I77371
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-395 <RES>
 A:Cross-references: GB:S66400; NID:9435697; PIDN:AAH27919.1; PID:9435700
 C:Genetics:
 A:Gene: GDB:CD44
 A:Cross-references: GDB:120739; OMIM:107269
 A:Map position: 11pter-11p13
 A:Introns: 257/1
 C:Superfamily: human cell adhesion protein CD44

Query Match 78.1%; Score 2036; DB 2; Length 395;
 Best Local Similarity 79.9%; Pred. No. 1,4e-123;
 Matches 394; Conservative 1; Mismatches 0; Indels 98; Gaps 1;

QY 1 MDKFWMAWAGLCVPLSLAQIDINTCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60
 |||||
 Db 1 MDKFWMAWAGLCVPLSLAQIDINTCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60
 QY 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120
 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120
 Db 61 PTMAQMERKALSIGETCRKYGFIKGVVAPRIHPNSICAAANTGYIILTSNTSOVDYICFN 120

QY 121 ASAPPEEDCTSVTLDPNAPDGPITITVNRDGRVYKGEYRTNPEDIPNSPTDDVSS 180
 |||
 DB 121 ASAPPEEDCTSVTLDPNAPDGPITITVNRDGRVYKGEYRTNPEDIPNSPTDDVSS 180
 |||
 QY 181 GSSSERSTSGGYLYFTFSTVHPIDEDSPWITDSTRIPATNMDSHSTTLOPTANPNT 240
 |||
 DB 181 GSSSERSTSGGYLYFTFSTVHPIDEDSPWITDSTRIPATNMDSHSTTLOPTANPNT 240
 |||
 QY 241 GLVEDLDRGTGLSMTTQGSNSQSFSTHBEGLBEKHPTTSTLTSSNRNDVTGGRDPNH 300
 |||
 DB 241 GLVEDLDRGTGLSMTTQGSNSQSFSTHBEGLBEKHPTTSTLTSSNRNDVTGGRDPNH 300
 |||
 QY 301 SEGSTTLLEGYTSHPHTKESRTEIPVTSKATGSGVAVTVGDSNSNVNRSLSGDQTF 360
 |||
 DB 258 -----DQDTF 262
 |||
 QY 361 HPSGGSHTTHGSESDGSHSGQEGGANTTSGPIRTPOIPFWLIIASLALALILAVCIA 420
 |||
 DB 263 HPSGGSHTTHGSESDGSHSGQEGGANTTSGPIRTPOIPFWLIIASLALALILAVCIA 420
 |||
 QY 421 VNSRRRCGQKKLVINGNGAVEDRKPSGLNGEASKSQEWVHLVKNSESTPPQFMTADE 480
 |||
 DB 323 VNSRRRCGQKKLVINGNGAVEDRKPSGLNGEASKSQEWVHLVKNSESTPPQFMTADE 382
 |||
 QY 481 TRNLQNVDMKIGV 493
 |||
 DB 383 TRNLQNVDMKIGV 395
 |||

RESULT 6

cell adhesion molecule CD44 - human
 C/Species: Homo sapiens (man)
 C/Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 01-Dec-2000
 C/Accession: JH0417; A32376; G02251; A32377
 R/Hart, H.J.; Isola, N.; Cooper, D.L.
 Biochem. Biophys. Res. Commun. 178, 1127-1134, 1991
 A/Title: The multispecific cell adhesion molecule CD44 is represented in reticulocyte
 A/Reference number: JH0417; MUID:91337049; PMID:1840467
 A/Accession: JH0417
 A/Molecule type: mRNA
 A/Residues: 1-361 <HAR>
 A/Cross-references: GB:M59040; NID:g180129; PIDN:AAA51950.1; PID:g180130
 A/Experimental source: reticulocyte
 A/Note: the authors translated the codon ATG for residues 63, 66 and 239 as Trp and GGA
 R/Stamenkovic, I.; Amiot, M.; Pesando, J.M.; Seed, B.
 Cell 56, 1057-1062, 1989
 A/Title: A lymphocyte molecule implicated in lymph node homing is a member of the cartl
 A/Reference number: A32376; MUID:89169434; PMID:2466575
 A/Accession: A32376
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-238, 'E', 240-361 <STA>
 A/Cross-references: GB:M24915; NID:g180196; PIDN:AAA3674.1; PID:g180197
 R/Bosch, P.P.; Stevens, J.W.; Buckwalter, J.A.; Matura, R.J.
 submitted to the EMBL Data Library, November 1995
 A/Reference number: H00921
 A/Accession: G02251
 A/Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: mRNA
 A/Residues: 1-25, 'W', 27-108, 'S', 110-361 <BOS>
 A/Cross-references: EMBL:U04373; NID:g101785; PID:g101786
 R/Goldstein, L.A.; Zhou, D.F.H.; Picker, L.D.; Minny, C.N.; Bargatze, R.F.; Ding, J.F.;
 Cell 56, 1063-1072, 1989
 A/Title: A human lymphocyte homing receptor, the herpes antigen, is related to cartilage
 A/Reference number: A32377; MUID:89169435; PMID:2466576
 A/Accession: A32377
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-108, 'S', 110-293, 'S', <GOL>
 A/Cross-references: GB:M5078; NID:g186660; PIDN:AAA5138.1; PID:g186661
 C/Superfamily: human cell adhesion protein CD44

C/Keywords: alternative splicing; cell adhesion; surface antigen; transmembrane prot
 F1269-286/Domain: transmembrane #status predicted <TM>

Query Match 70.4%; Score 1835; DB 2; Length 361;
 Best Local Similarity 72.8%; Pred. No. 1e-110;
 Matches 359; Conservative 0; Mismatches 2; Indels 132; Gaps 1;

QY 1 MDKFWTHAAMGLCVPLSLAIDINICRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
 |||
 DB 1 MDKFWTHAAMGLCVPLSLAIDINICRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
 |||
 QY 61 PTMAQMERALSIGFETCRVGFIEGHVAVPRIHPNSICAAANTGYIILTYNTSOYDTYCFN 120
 |||
 DB 61 PTMAQMERALSIGFETCRVGFIEGHVAVPRIHPNSICAAANTGYIILTYNTSOYDTYCFN 120
 |||
 QY 121 ASAPPEEDCTSVTLDPNAPDGPITITVNRDGRVYKGEYRTNPEDIPNSPTDDVSS 180
 |||
 DB 121 ASAPPEEDCTSVTLDPNAPDGPITITVNRDGRVYKGEYRTNPEDIPNSPTDDVSS 180
 |||
 QY 181 GSSSERSTSGGYLYFTFSTVHPIDEDSPWITDSTRIPATNMDSHSTTLOPTANPNT 240
 |||
 DB 181 GSSSERSTSGGYLYFTFSTVHPIDEDSPWITDSTRIPATNMDSHSTTLOPTANPNT 240
 |||
 QY 241 GLVEDLDRGTGLSMTTQGSNSQSFSTHBEGLBEKHPTTSTLTSSNRNDVTGGRDPNH 300
 |||
 DB 224 ----- 223
 |||
 QY 301 SEGSTTLLEGYTSHPHTKESRTEIPVTSKATGSGVAVTVGDSNSNVNRSLSGDQTF 360
 |||
 DB 224 -----DQDTF 228
 |||
 QY 361 HPSGGSHTTHGSESDGSHSGQEGGANTTSGPIRTPOIPFWLIIASLALALILAVCIA 420
 |||
 DB 229 HPSGGSHTTHGSESDGSHSGQEGGANTTSGPIRTPOIPFWLIIASLALALILAVCIA 288
 |||
 QY 421 VNSRRRCGQKKLVINGNGAVEDRKPSGLNGEASKSQEWVHLVKNSESTPPQFMTADE 480
 |||
 DB 289 VNSRRRCGQKKLVINGNGAVEDRKPSGLNGEASKSQEWVHLVKNSESTPPQFMTADE 348
 |||
 QY 481 TRNLQNVDMKIGV 493
 |||
 DB 349 TRNLQNVDMKIGV 361
 |||

RESULT 7

A30901
 lymphocyte adhesion receptor precursor - baboon
 C/Species: Papio sp. (baboon)
 C/Date: 18-Apr-1989 #sequence_revision 18-Apr-1989 #text_change 21-Jul-2000
 C/Accession: A33935; A30901
 R/Ridzenda, R.L.; Carter, W.G.; Nottenburg, C.; Wayner, E.A.; Gallatin, W.M.; St. Jo
 Proc. Natl. Acad. Sci. U.S.A. 86, 4659-4663, 1989
 A/Title: Isolation and DNA sequence of a cDNA clone encoding a lymphocyte adhesion
 A/Reference number: A33935; MUID:89282830; PMID:2411974
 A/Accession: A33935
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-362 <IDZ>
 A/Cross-references: GB:M22452; NID:g176576; PID:g176577
 C/Comment: This protein was isolated from the herpes papio induced B cell lymphoma.
 C/Genetics:
 A/Name: CD44; ECRH11; Herpes-1 antigen
 C/Superfamily: human cell adhesion protein CD44
 C/Keywords: cytoskeleton; extracellular matrix; glycoprotein; lipoprotein; surface
 F1-20/Domain: signal sequence #status predicted <SIG>
 F1-21-62/Product: lymphocyte adhesion receptor #status predicted <MNT>
 F1-21-62/Product: lymphocyte adhesion receptor #status predicted <MNT>
 F1-271-290/Domain: transmembrane #status predicted <TM>
 F1-291-362/Domain: intracellular #status predicted <CYT>
 F1-25,57,100,110,120,256/Binding site: carbohydrate (Asn) (covalent) #status predicte
 F1-296/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 65.1%; Score 1698.5; DB 2; Length 362;

Best Local Similarity 68.2%; Pred. No. 6,1e-102; Indels 133; Gaps 2;
Matches 337; Conservative 7; Mismatches 17;

```

QY 1 MDKFWMAAGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTL 60
DB 1 MDKFWMAAGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTL 60
QY 61 PTMAQMEKALSIGETCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCRN 120
DB 61 PTMAQMEKALSIGETCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCRN 120
QY 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180
DB 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180
QY 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180
DB 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180
QY 181 GSSSERSTSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANRN 239
DB 181 GSSSERSTSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANRN 239
QY 240 TGLVEDDRTGPLSMITQGSNSQSFSTSHGLEEDKDHPTTSTLSSNRNDVTGGRDEN 299
DB 240 TGLVEDDRTGPLSMITQGSNSQSFSTSHGLEEDKDHPTTSTLSSNRNDVTGGRDEN 299
QY 225 ----- 224
DB 225 ----- 224
QY 300 HSEGSTLLSGYISHYHTKESRTFIPVTSAGTGSFGVTAVTVGDSNSVNRSLSGDQT 359
DB 225 ----- 224
QY 360 FHPEGSHTHGSDSHSGOEGGANTTSGPIRTPOIPEWLIILASLALALILAVCI 419
DB 229 FDBGSGHTTHGSSASHSGSRGANTTSGPIRTPOIPEWLIILASLALALILAVCI 288
QY 420 AVNSRRRCGQKKKLVINGAVEDRPSGLNGEASKSQEMVHLVNKESSETPDQFWTA 479
DB 289 AVNSRRRCGQKKKLVINGAVEDRPSGLNGEASKSQEMVHLVNKESSETPDQFWTA 348
QY 480 ETRRLQNVDMKIGV 493
DB 349 ETRRLQNVDMKIGV 362

```

RESULT 8

224240
Lymphocyte surface antigen CD44 precursor - horse
C:Species: Equus caballus (domestic horse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000
C/Accession: I46245; S24240
R/Tavernor, A.S.; Deyerson, E.V.; Coadwell, W.J.; Lunn, D.P.; Zhang, C.; Davis, W.; Butc
Immunogenetics 37, 474-477, 1993
A>Title: Molecular cloning of equine CD44 cDNA by a COS cell expression system.
A/Reference number: I46245; MUID:93170897; PMID:8436424
A/Accession: I46245
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-359 <T12>
A/Cross-references: EMBL:X66862; NID:91059; PIDN:CAA47331.1; PID:91060
C/Superfamily: human cell adhesion protein CD44
C/Keywords: surface antigen; transmembrane protein

Query Match 59.3%; Score 1547; DB 2; Length 359;
Best Local Similarity 63.0%; Pred. No. 3.3e-92; Indels 138; Gaps 5;
Matches 312; Conservative 20; Mismatches 25;

```

QY 1 MDKFWMAAGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTL 60
DB 1 MDKFWMAAGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTL 60
QY 61 PTMAQMEKALSIGETCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCRN 120
DB 61 PTMAQMEKALSIGETCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCRN 120
QY 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180
DB 121 ASAPPEEDCTSVTDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSS 180

```

QY 181 GSSSERSTSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANRN 238

DB 181 GSSSER-STSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANRN 223

QY 239 NTGLVEDDRTGPLSMITQGSNSQSFSTSHGLEEDKDHPTTSTLSSNRNDVTGGRDP 298

DB 224 ----- 231

QY 239 NNSGSTLLSGYISHYHTKESRTFIPVTSAGTGSFGVTAVTVGDSNSVNRSLSGDQT 358

DB 232 ----- 231

QY 359 FHPEGSHTHGSDSHSGOEGGANTTSGPIRTPOIPEWLIILASLALALILAVCI 418

DB 232 ----- 284

QY 419 IAVNSRRRCGQKKKLVINGAVEDRPSGLNGEASKSQEMVHLVNKESSETPDQFWTA 478

DB 285 IAVNSRRRCGQKKKLVINGAVEDRPSGLNGEASKSQEMVHLVNKESSETPDQFWTA 344

QY 479 ETRRLQNVDMKIGV 493

DB 345 ETRRLQNVDMKIGV 359

RESULT 9

545305
CD44 antigen precursor - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C/Date: 20-Oct-1994 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C/Accession: S45305
R/Milde, K.F.; Alejandro, R.; Mintz, D.H.; Pastori, R.L.
Biochem. Biophys. Acta 1218, 112-114, 1994
A>Title: Molecular cloning of the canine CD44 antigen cDNA.
A/Reference number: S45305; MUID:94250687; PMID:7514890
A/Accession: S45305
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-351 <M12>
A/Cross-references: EMBL:Z27115; NID:9473226; PIDN:CAA61630.1; PID:9473227
C/Superfamily: human cell adhesion protein CD44

Query Match 56.4%; Score 1470; DB 2; Length 351;
Best Local Similarity 61.7%; Pred. No. 2.9e-87; Indels 132; Gaps 3;
Matches 297; Conservative 21; Mismatches 31;

```

QY 9 AMGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTLPTMAQMEK 68
DB 2 AMGLCLVPLSLAIDLNITCRFAGVFEHKNRGYSISRTAADLCKAFNSTLPTMAQMEK 61
QY 69 ALSTGFEFCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCFNASAPPEED 128
DB 62 ALSTGFEFCRGFIIEGHVYIPRIHPNSICANNNGVYILTSNTSOYDTCFNASAPPEED 121
QY 129 CTSVVDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSSGSSSERSS 188
DB 122 CTSVVDLPAFAFDGPITITIVNRDGRYVYQGEYRNPEDIPSPETDDVSSGSSSERSS 181
QY 189 TSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANNTGLVELD 247
DB 182 TSGGYIFFT-FSTVHPIPEDSPWITDSDRIIPATNMDSHSTLQPTANNTGLVELD 216
QY 248 RTGPLSMITQGSNSQSFSTSHGLEEDKDHPTTSTLSSNRNDVTGGRDENHSGSTTL 307
DB 217 ----- 219
QY 308 LEGYTSHYPTKESRTFIPVTSAGTGSFGVTAVTVGDSNSVNRSLSGDQTFHPSGSH 367
DB 220 ----- 230
QY 368 TTHGSDSHSGOEGGANTTSGPIRTPOIPEWLIILASLALALILAVCIANNSRRRC 427

```

Db 231 TTHSGSAGHSSGQEGGANTTSGPMKRPQIPMLIIASLALALILAVCIANVSRRC 290

Qy 428 GQKKYVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPPQMTADETRLQNV 487

Db 291 GQKKYVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPPQMTADETRLQNV 350

Qy 488 D 488

Db 351 D 351

RESULT 10

A53286
cell-surface glycoprotein CD44 precursor - bovine
N:Alternate names: CD44 protein
C:Species: Bos primigenius taurus (cattle)
C>Date: 02-May-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A53286; #22123
R:Boeworth, B.T.; St John, T.; Gallatin, W.M.; Harp, J.A.
Mol. Immunol. 28, 1131-1135, 1991
A:Title: Sequence of the bovine CD44 cDNA: comparison with human and mouse sequences.
A:Reference number: A53286; MUID:92017904; PMID:11922105
A:Accession: A53286
A:Molecule type: mRNA
A:Residues: 1-366 <BOS>
A:Cross-references: EMBL:X62881; NID:g186; PID:CA444675.1; PID:g187
A:Note: sequence extracted from NCBI backbone (NCBI:63418, NCBI:63419)
C:Superfamily: human cell adhesion protein CD44
C:Keywords: cell adhesion; glycoprotein; transmembrane protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-366/Product: cell-surface glycoprotein CD44 #status predicted <MAT>
F:274-294/Domain: transmembrane #status predicted <TM>
F:25,57,100,110,120,222,260/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 56.3%; Score 1469.5; DB 2; Length 366;
Best Local Similarity 61.0%; Pred. No. 3.2e-87;
Matches 305; Conservative 22; Mismatches 32; Indels 141; Gaps 8;

Qy 1 MDKFWHAAWGLCLVPLSLAQIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60

Db 1 MDKFWHAAWGLCLVPLSLAQIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60

Qy 61 PTMAQMEKALSIGFETCRGFIIEGHVYIRIHPNSICANNNGVYIITSTSYDYTCEN 120

Db 61 PTMAQMEKALSIGFETCRGFIIEGHVYIRIHPNSICANNNGVYIITSTSYDYTCEN 120

Qy 121 ASAPPEEDCTSVTLDPNAFDPGPIITIVNRDGRVYKGEYRTNPDIDYPSNPTDDV 175

Db 121 ASAPPEEDCTSVTLDPNAFDPGPIITIVNRDGRVYKGEYRTNPDIDYPSNPTDDV 180

Qy 176 DVVSGSSERSSTISGCI-IFVT-FSTVHP-IPDEDSFWITDSTRIPATNMDSSHSTLLQ 223

Db 181 DVVSGSSERSSTISGCI-IFVT-FSTVHP-IPDEDSFWITDSTRIPATNMDSSHSTLLQ 223

Qy 224 PTANPTGLVEDLDRTPGLSMITQOQNSQSFSTHEGLDEKHPHTSTLTSNRNDVYG 223

Db 224 PTANPTGLVEDLDRTPGLSMITQOQNSQSFSTHEGLDEKHPHTSTLTSNRNDVYG 224

Qy 294 GRDPNHSSEGSTLLLEGYTSHPHTKESRTFIPVTSKATGSGFVAVTAVGDSNVNRS 353

Db 294 GRDPNHSSEGSTLLLEGYTSHPHTKESRTFIPVTSKATGSGFVAVTAVGDSNVNRS 353

Qy 354 SGDDTFHPSGSHRTTHSGSDGSHSGQEGGANTTSGPIRTPOIPMLIIASLALAL 413

Db 354 SGDDTFHPSGSHRTTHSGSDGSHSGQEGGANTTSGPIRTPOIPMLIIASLALAL 413

Qy 414 ILAVCIANVSRRCQKKKLVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPD 473

Db 414 ILAVCIANVSRRCQKKKLVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPD 473

Qy 474 QPMTADETRLQNVDMKIGV 493

Db 474 QPMTADETRLQNVDMKIGV 493

RESULT 11

A35616
T-cell surface glycoprotein CD44 - hamster
C:Species: Crictetinae gen. sp. (hamster)
C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 21-Jul-2000
C:Accession: A35616
R:Aruffo, A.; Stamenkovic, I.; Melnick, M.; Underhill, C.B.; Seed, B.
Cell 61, 1303-1313, 1990
A:Title: CD44 is the principal cell surface receptor for hyaluronate.
A:Reference number: A35616; MUID:90304889; PMID:1694723
A:Accession: A35616
A:Molecule type: mRNA
A:Residues: 1-362 <ARU>
C:Superfamily: human cell adhesion protein CD44
C:Keywords: cell adhesion; glycoprotein; transmembrane protein

Query Match 54.7%; Score 1427.5; DB 2; Length 362;
Best Local Similarity 59.6%; Pred. No. 1.6e-84;
Matches 297; Conservative 23; Mismatches 37; Indels 141; Gaps 7;

Qy 1 MDKFWHAAWGLCLVPLSLAQIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNS 58

Db 1 MDKFWHAAWGLCLVPLSLAQIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNS 60

Qy 59 TLPTMAQMEKALSIGFETCRGFIIEGHVYIRIHPNSICANNNGVYIITSTSYDYTC 118

Db 59 TLPTMAQMEKALSIGFETCRGFIIEGHVYIRIHPNSICANNNGVYIITSTSYDYTC 120

Qy 119 FNASAPPEEDCTSVTLDPNAFDPGPIITIVNRDGRVYKGEYRTNPDIDYPSNPTDDV 178

Db 121 FNASAPPEEDCTSVTLDPNAFDPGPIITIVNRDGRVYKGEYRTNPDIDYPSNPTDDV 180

Qy 179 SSGSSERSSTISGCI-IFVT-FSTVHP-IPDEDSFWITDSTRIPATNMDSSHSTLLQPTAN 237

Db 181 SSGSSERSSTISGCI-IFVT-FSTVHP-IPDEDSFWITDSTRIPATNMDSSHSTLLQPTAN 218

Qy 228 PVTGLVEDLDRTPGLSMITQOQNSQSFSTHEGLDEKHPHTSTLTSNRNDVYGGRD 297

Db 228 PVTGLVEDLDRTPGLSMITQOQNSQSFSTHEGLDEKHPHTSTLTSNRNDVYGGRD 222

Qy 229 PNHSEGSTLLLEGYTSHPHTKESRTFIPVTSKATGSGFVAVTAVGDSNVNRSLSGDQ 357

Db 229 PNHSEGSTLLLEGYTSHPHTKESRTFIPVTSKATGSGFVAVTAVGDSNVNRSLSGDQ 224

Qy 358 D-TFHPSGSHRT-THSGSDGSHSGQEGGANTTSGPIRTPOIPMLIIASLALAL 415

Db 358 D-TFHPSGSHRT-THSGSDGSHSGQEGGANTTSGPIRTPOIPMLIIASLALAL 415

Qy 416 AVCIANVSRRCQKKKLVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPDQ 475

Db 416 AVCIANVSRRCQKKKLVINGNGAVNEVRKPSGLNGEASKSQEMVHLVNTKSSSTPDQ 475

Qy 476 QPMTADETRLQNVDMKIGV 493

Db 476 QPMTADETRLQNVDMKIGV 362

RESULT 12

B38745
cell adhesion molecule CD44 precursor, long form (meta-1) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 05-Nov-1999
C:Accession: B38745; A38745
R:Guenther, U.; Hofmann, M.; Rudy, W.; Reber, S.; Zoeller, M.; Haubmann, I.; Matzke
Cell 65, 13-24, 1991
A:Title: A new variant of glycoprotein CD44 confers metastatic potential to rat car

A:Reference number: A38745; MUID:91191552; PMID:1707342
A:Accession: B38745
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-503 <GU>
 A:Cross-references: GB:M61874; NID:g576534; PIDN:AAA53534.1; PID:g576535
 A:Accession: A38745
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-223,386-503 <GU2>
 A:Cross-references: GB:M61875
 C:Keywords: cell adhesion

Query Match 54.6%; Score 1425; DB 2; Length 503;
 Best Local Similarity 58.7%; Pred. No. 3.6e-84;
 Matches 315; Conservative 36; Mismatches 108; Indels 78; Gaps 16;

```

Qy 1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 57
Db 1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 60
Qy 58 STLEPTMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSQYDT 116
Db 61 TLLEPTMAQMEKALRKGEFCRCYGFIEGHVVIPIRHNAICANNNGVYIILASNTSHYDT 120
Qy 117 YCFNAPAPPEEDCTSVTDLPNAPFGPITITVNRDGRYQKGYRTNPEIDVSPNPTD 176
Db 121 YCFNAPAPPEEDCTSVTDLPNAPFGPITITVNRDGRYQKGYRTNPEIDVSPNPTD 180
Qy 177 DVSSGSSERSSTGGYIFYT-PSTVPIPEDD-----SPWITDST---D 217
Db 181 DVSSGSSERSSTGGYIFYT-PSTVPIPEDD-----SPWITDST---D 217
Qy 218 RIRPTNMDSSHSTLQCTANPNGLVEDLDR--TGPLSMTQOOSQSPSTHGLEHOK 275
Db 240 RIRPTNMDSSHSTLQCTANPNGLVEDLDR--TGPLSMTQOOSQSPSTHGLEHOK 275
Qy 276 DHPTSTLTSSNRNDVYG-----GRDPNHSSESTLLLEGYT--SHYPTKE 320
Db 295 PHATSTWADPNSTTEBAATQKEKFEENWQKMPPTSEDS-HVTEGTTASANNHPQ 353
Qy 321 SRPIPTSAKTSFGYATVAVGDSNNVRSLSGDODT---HSGSGHTHGESBQ 376
Db 354 RMT-----TQSCEDVSMT-----DFPDPSHPWGGQHOT--ESKG 386
Qy 377 HSHSGDEGANTTSGPIRTPOIPEWIIILASLALAILAVCIANVSRRCQKKLVIN 436
Db 387 HSHSGDEGANTTSGPIRTPOIPEWIIILASLALAILAVCIANVSRRCQKKLVIN 446
Qy 437 SGNGAVDRKPSGLNGASYSQEMVHLVKNESSETPQENTADETRLQNVDMKIGV 493
Db 447 SGNGAVDRKPSGLNGASYSQEMVHLVKNESSETPQENTADETRLQNVDMKIGV 503

```

RESULT 13
 A37009
 CD44 homolog membrane glycoprotein precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 21-Jul-2000
 A:Accession: A37009
 R/Zhou, D.F.H.; Ding, J.F.; Pickler, L.J.; Bargatzke, R.F.; Butcher, E.C.; Goeddel, D.V.
 J. Immunol. 143, 3390-3395, 1989
 A:Title: Molecular cloning and expression of Bgp-1, The mouse homolog of the human H-CAM
 A:Reference number: A37009; MUID:90038499; PMID:2681416
 A:Accession: A37009
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-363 <ZHO>
 A:Cross-references: GB:M30655; NID:g200332; PIDN:AAA3922.1; PID:g200333
 C:Superfamily: human cell adhesion protein CD44
 C:Keywords: cell adhesion; glycoprotein; membrane protein

Query Match 52.5%; Score 1368; DB 2; Length 363;
 Best Local Similarity 58.1%; Pred. No. 1.1e-80;
 Matches 289; Conservative 28; Mismatches 42; Indels 138; Gaps 9;

1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 58

```

Db 1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 60
Qy 59 TLPTMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSQYDT 117
Db 61 TLPTMDQKALSKGFECRCYGFIEGNVVIPIRHNAICANNNGVYIILASNTSHYDT 120
Qy 118 CFNAPAPPEEDCTSVTDLPNAPFGPITITVNRDGRYQKGYRTNPEIDVSPNPTD 177
Db 121 CFNAPAPPEEDCTSVTDLPNAPFGPITITVNRDGRYQKGYRTNPEIDVSPNPTD 180
Qy 178 VSSGSSERSSTGGYIFYT-PSTVPIPEDD-----SPWITDST---D 217
Db 181 VSSGSSERSSTGGYIFYT-PSTVPIPEDD-----SPWITDST---D 217
Qy 238 PNTGLVEDLDRGTPLSMTQOOSQSPSTHGLEHOKHTTSTLTSSNRNDVYGARD 297
Db 209 -----DQDSFFIRSTLARDSDS----- 227
Qy 298 PNHSEGSTLLLEGYTSHYPTKESTRTPIPTSAKTSFGYATVAVGDSNNVRSLSGDQ 357
Db 228 -----SKDSR----- 232
Qy 358 DTFHPSGSSHT-TGSSSDGSHSGDEGANTTSGPIRTPOIPEWIIILASLALAILA 416
Db 233 -----GSSRTVTGSELGHSANQDSGVTTSFGPMRPOIPEWIIILASLALAILA 286
Qy 417 VCIANVSRRCQKKLVINGANGAVEDEKPSGLNGASYSQEMVHLVKNESSETPQFM 476
Db 287 VCIANVSRRCQKKLVINGANGAVEDEKPSGLNGASYSQEMVHLVKNESSETPQFM 346
Qy 477 TADETRNLQNVDMKIGV 493
Db 347 TADETRNLQNVDMKIGV 363

```

RESULT 14
 A34424
 Cd44 membrane glycoprotein precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 01-Dec-2000
 A:Accession: A34424; A34907
 R/Nottebohm, C.; Rees, G.; St. John, T.
 Proc. Natl. Acad. Sci. U.S.A. 86, 8521-8525, 1989
 A:Title: Isolation of mouse Cd44 cDNA: structural features are distinct from the pri
 A:Reference number: A34424; MUID:90046829; PMID:2682651
 A:Accession: A34424
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-365 <NOT>
 A:Cross-references: GB:M27130; NID:G192530; PIDN:AAA37407.1; PID:g309161
 R/Wolfe, E.U.; Gause, W.C.; Pelfrey, C.M.; Holland, S.M.; Steinberg, A.D.; August,
 U. Biol. Chem. 265, 341-347, 1990
 A:Title: The cDNA sequence of mouse Bgp-1 and homology to human CD44 cell surface ar
 A:Reference number: A34907; MUID:90044420; PMID:2403559
 A:Accession: A34907
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 8-195, 'G', 197-365 <MO>
 A:Cross-references: GB:J05163; NID:g200334; PIDN:AAA3923.1; PID:g200335
 C:Superfamily: human cell adhesion protein CD44
 C:Keywords: cell adhesion; glycoprotein; membrane protein

Query Match 52.2%; Score 1361; DB 2; Length 365;
 Best Local Similarity 57.7%; Pred. No. 3e-80;
 Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

```

Qy 1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAF 56
Db 1 MDKFWMAAMG-LCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAF 60
Qy 57 NSTLEPTMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSQYD 115
Db 57 NSTLEPTMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSQYD 115

```

```
Db 61 NSTLPTMDQKTLAKSGFETCRVGFIEGNVIRIHPNACANHTGVILVTSNTSHYD 120
QY 116 TYCFNAGAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDIYPSNPTD 175
Db 121 TYCFNAGAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDIYPSNPTD 180
QY 176 DDVSSGSSSSRSSTSGVIFPTFSTVHIPDEDSPTITDSDRIPATNMDSSHSTLQPT 235
Db 181 DDVSSGSTIEK-STPESTYIAHTY-----LPT-----QPT 209
QY 236 ANPNTGLVEDLDRTGPLSMTTQGSNSQSFSTSHGLEEDKDHPSTLTSSNNDVTGGR 295
Db 210 G-----DQDSFFIRSTLIATRDSD----- 229
QY 296 RDPNHSSESTLLBGTSHYPTKESRTPFIVTSAKTSGFVTAATVGDNSNVNRSLSG 355
Db 230 -----SKDSR----- 234
QY 356 DQDTFHPSGGSHT-THGSESDGSHSGQEGGANTTSGPIRTPQIPEWLIILASLALALI 414
Db 235 -----GSRRTVTHGSELHGHSSANDSGVTTTSGPMRRPQIPEWLIILASLALALI 286
QY 415 LAVCIANVRRRCQKKLVNNGAVEDRKPSGLNGEASKSGQEMVHLVNXKESSETPDQ 474
Db 287 LAVCIANVRRRCQKKLVNNGAVEDRKPSGLNGEASKSGQEMVHLVNXKESSETPDQ 346
QY 475 FMTADETNLQNVDMKIGV 493
Db 347 CMTADETNLQNVDMKIGV 365
```

RESULT 15

```
S24222
CD4 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: S24222
R:Hotmann, M.; Rudy, W.; Zoeller, M.; Toeig, C.; Ponta, H.; Herrlich, P.; Guenther, U.
Cancer Res. 51, 5292-5297, 1991
A:Title: CD4 splice variants confer metastatic behavior in rats: homologous sequences a
A:Reference number: S24222; MUID:92005448; PMID:1717145
A:Accession: S24222
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-337 <HOB>
A:Cross-references: EMBL:X62739; NID:937651; PIDN:CAA44602.1; PID:937652
```

Query Match 26.8%; Score 699; DB 2; Length 337;

Best local similarity 69.7%; Pred. No. 9.6e-38;

Matches 147; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

```
QY 161 YRTNPEDIYPSNPTDDVSSGSSRSSTSGVIFYFSTVHPF-----PDEDSPTITD 214
Db 146 YRQTPRE-----DSHSTGTGAASATTS-----HPQGRITTPSPEDSN-ITD 186
QY 215 STDRIP-----ATNMDSSHSTLQPTANPNTGLVEDLDRTGPLSMTTQGSNSQS 263
Db 187 FFNPI SHPMGRGHQAGRMDWDSHSTLQPTANPNTGLVEDLDRTGPLSMTTQGSNSQS 246
QY 264 FSTSHGLEEDKDHPSTLTSSNNDVTGGRDPNHSSESTLLBGTSHYPTKESRT 323
Db 247 FSTSHGLEEDKDHPSTLTSSNNDVTGGRDPNHSSESTLLBGTSHYPTKESRT 306
QY 324 FIPVTSKATGSGFVTAATVGDNSNVNRSLS 354
Db 307 FIPVTSKATGSGFVTAATVGDNSNVNRSLS 337
```

Search completed: March 8, 2004, 06:12:10
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 / Search time 17 Seconds

(without alignments)
1510.034 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 2608
Sequence: 1 MDKFWHAAAGLCLVPLSLA.....QFMTADETRNLQNVDMKGV 493

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2468.5	94.7	742	1 CD44_HUMAN	P16070 h cd44 anti
2	1698.5	65.1	362	1 CD44_PAPHA	P14745 papio hamad
3	1662.5	63.7	778	1 CD44_MOUSE	P15379 mus musculu
4	1631	62.5	431	1 CD44_MESAU	Q60522 m cd44 anti
5	1547	59.3	359	1 CD44_HORSE	Q05078 equus cabal
6	1470	56.4	351	1 CD44_CANPA	Q28284 canis famli
7	1469.5	56.3	366	1 CD44_BOVIN	Q29423 bos taurus
8	1425	54.6	503	1 CD44_RAT	P26051 rattus norv
9	1398.5	53.6	362	1 CD44_CRIGR	P20944 cricetulus
10	1377	6.8	389	1 SER1_BOMMO	P07856 bombyx mori
11	171	6.6	276	1 TS66_RABIT	P98065 oryctolagus
12	168	6.4	2124	1 PGCA_RAT	P07897 rattus norv
13	163	6.2	275	1 TS66_MOUSE	Q08859 mus musculu
14	162.5	6.2	277	1 TS66_HUMAN	Q02496 mus musculu
15	158	6.1	630	1 MUC1_MOUSE	Q6182 mus musculu
16	155.5	6.0	2132	1 PGCV_MOUSE	Q62059 mus musculu
17	154	5.9	3358	1 PGCV_MOUSE	Q62059 mus musculu
18	152.5	5.8	3381	1 PGCV_BOVIN	P81882 bos taurus
19	151.5	5.8	2333	1 PGCA_CANPA	Q28343 canis famli
20	151	5.8	2109	1 PGCA_HUMAN	P07898 gallus galli
21	150.5	5.8	636	1 YN66_CHICK	P53882 saccharomyc
22	150	5.8	862	1 PGCV_MOUSE	Q28858 macaca neme
23	148	5.7	1357	1 AMYR_YEAST	P08640 saccharomyc
24	147	5.6	3178	1 YS89_CAEEL	Q09624 caenorhabdi
25	146	5.6	2415	1 PGCA_HUMAN	P16112 homo sapien
26	145	5.6	883	1 PGCB_RAT	P55068 rattus norv
27	143.5	5.5	1306	1 MSB2_YEAST	P32334 saccharomyc
28	141.5	5.4	937	1 HYR1_CANAL	P46591 candida alb
29	141	5.4	797	1 VGLX_HSVB	Q28668 equine hepr
30	139	5.3	537	1 PGCA_PIG	Q29011 sus scrofa
31	139	5.3	681	1 VGF_MABVP	P35254 martus vir
32	139	5.3	1161	1 DAM4_YEAST	P47179 saccharomyc
33	139	5.3	1260	1 AL51_CANAL	P46590 candida alb

34	139	5.3	3396	1 PGCV_HUMAN	P13611 homo sapien
35	137	5.3	1169	1 YK82_YEAST	P36170 saccharomyc
36	136.5	5.2	1253	1 DSPF_HUMAN	Q09104 homo sapien
37	135	5.2	907	1 VGP3_EBV	P03200 Epstein-Barr
38	134	5.1	3562	1 PGCV_CHICK	Q90953 gallus galli
39	133.5	5.1	556	1 MSC3_YEAST	Q12215 saccharomyc
40	133.5	5.1	2316	1 PTP2_RAT	Q62656 rattus norv
41	133.5	5.1	670	1 YFEG_SCHPO	Q13854 schistosach
42	133	5.1	1257	1 PGCV_RAT	P55067 rattus norv
43	133	5.1	5703	1 MUS8_HUMAN	P50677 homo sapien
44	132.5	5.1	1403	1 YDF3_SCHPO	Q10475 schistosach
45	132	5.1	1075	1 FLOS_YEAST	P38994 saccharomyc

ALIGNMENTS

RESULT 1
CD44_HUMAN STANDARD, PRT, 742 AA.
ID CD44_HUMAN
AC P16070; P22511; Q04858; Q13419; Q13957; Q13958; Q13959; Q13960;
AC Q13961; Q13967; Q13968; Q13980; Q15861; Q16064; Q16065; Q16066;
AC Q16208; Q16522; Q96024;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CD44 antigen precursor (Phagocytic glycoprotein I) (PGP-I) (HUTCH-I)
DE (Extracellular matrix receptor-II) (ECMR-II) (GP90 lymphocyte
DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)
DE (Heparan sulfate proteoglycan) (Epicam) (CDW44).
GN CD44 OR IHR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Lymphoblast;
RX MEDLINE=93101687; PubMed=1465456;
RA Screation G.R., Bell M.V., Jackson D.G., Cornelis F.B., Gerth U.,
RA Bell U.I.;
RT "Genomic structure of DNA encoding the lymphocyte homing receptor
RT CD44 reveals at least 12 alternatively spliced exons.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM EPIDERMAL).
RC TISSUE=Keratinocytes;
RX Kugelman L.C., Ganguly S., Haggerty J.G., Weissman S.M.,
RA Milstone L.M.;
RT "The core protein of epicam, a heparan sulfate proteoglycan on
RT keratinocytes, is an alternative form of CD44.";
RL J. Invest. Dermatol. 99:381-385(1992).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM EPITHELIAL CD44E).
RC MEDLINE=91122041; PubMed=1991450;
RX Stemekovic I., Aruffo A., Amiot M., Seed B.;
RA "The hematopoietic and epithelial forms of CD44 are distinct
RT polypeptides with different adhesion potentials for
RT hyaluronate-bearing cells.";
RL EMBO J. 10:343-348(1991).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS CD44R1 AND CD44R2).
RC TISSUE=Myeloid leukemia cells;
RX MEDLINE=91277598; PubMed=2056274;
RA Dougherty G.J., Lansdorp P.M., Cooper D.L., Humphries R.K.;
RT "Molecular cloning of CD44R1 and CD44R2, two novel isoforms of the
RT human CD44 lymphocyte 'homing' receptor expressed by hemopoietic
RT cells.";
RL J. Exp. Med. 174:1-5(1991).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS CD44SF, CD44R4 AND CD44R5).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93356912; PubMed=8352881;

RA Tanabe K.K., Nishi T., Sava H.;
 RT "Novel variants of CD44 arising from alternative splicing: changes in
 RT the CD44 alternative splicing pattern of MCF-7 breast carcinoma cells
 RT treated with hyaluronidase.";
 RT Mol. Carcinog. 7:212-220(1993).
 RL [6]
 RN
 RP SEQUENCE FROM N.A. (ISOFORM RETICULOCYTE).
 RC TISSUE=Reticulocytes;
 RX MEDLINE=91337049; PubMed=1840487;
 RT Hahn H.J., Isola N., Cooper D.L.,
 RT "The multispecific cell adhesion molecule CD44 is represented in
 RT reticulocyte CDNA.";
 RL Biochem. Biophys. Res. Commun. 178:1127-1134(1991).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89168434; PubMed=2466575;
 RT Stamenkovic I., Amiot M., Pesando J.M., Seed B.;
 RT "A lymphocyte molecule implicated in lymph node homing is a member of
 RT the cartilage link protein family.";
 RL Cell 56:1057-1062(1989).
 RN [8]
 RP SEQUENCE FROM N.A. (ISOFORMS SHORT-TAILED HEMATOPOIETIC AND CD44H).
 RX MEDLINE=89168435; PubMed=2466576;
 RA Goldstein L.A., Zhou D.F.H., Plicker L.J., Minty C.N., Bargatze R.F.,
 RA Ding J.F., Butcher E.C.;
 RT "A human lymphocyte homing receptor, the Hermes antigen, is related
 RT to cartilage proteoglycan core and link proteins.";
 RL Cell 56:1063-1072(1989).
 RN [9]
 RP SEQUENCE FROM N.A. (ISOFORM WITHOUT EXON 6).
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toehlyuk S., Carninci P., Prange C.,
 RA Rana S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gamarale P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Paley J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schercher A., Schein J.E., Jones S.J.W., Marra W.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [10]
 RP SEQUENCE OF 184-222 AND 473-625 FROM N.A. (ISOFORM KERATINOCYTE).
 RC TISSUE=Forelimb;
 RX MEDLINE=91177958; PubMed=2007624;
 RA Brown T.A., Bouchard T., St John T., Wayne E., Carter W.G.;
 RT "Human keratinocytes express a new CD44 core protein (CD44S) as a
 RT heparan-sulfate intrinsic membrane proteoglycan with additional
 RT exons.";
 RL J. Cell Biol. 113:207-221(1991).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Articular cartilage;
 RX Boshch P.P., Stevens J.W., Buckwalter J.A., Midura R.J.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE OF 267-603 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=92005448; PubMed=1171745;
 RA Hofmann M., Rudy W., Zoeller M., Toelg C., Ponta H., Herrlich P.,
 RA Guenther U.;
 RT "CD44 splice variants confer metastatic behavior in rats: homologous

RT sequences are expressed in human tumor cell lines.";
 RL Cancer Res. 51:5292-5297(1991).
 RN [13]
 RP SEQUENCE OF 223-265 FROM N.A.
 RX MEDLINE=94198700; PubMed=8146709;
 RA Matsumura Y., Hanbury D., Smith J., Tatin D.;
 RT "Non-invasive detection of malignancy by identification of unusual
 RT CD44 gene activity in exfoliated cancer cells.";
 RL BMJ 308:619-624(1994).
 RN [14]
 RP SEQUENCE OF 1-22 FROM N.A.
 RC TISSUE=Lymphoblast;
 RX MEDLINE=92017823; PubMed=1922057;
 RA Shivelman E., Bishop J.M.;
 RT "Expression of CD44 is repressed in neuroblastoma cells.";
 RL Mol. Cell. Biol. 11:5446-5453(1991).
 RN [15]
 RP REVIEW ON FUNCTION AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE=22399881; PubMed=12511867;
 RA Ponta H., Sherman L., Herrlich P.A.;
 RT "CD44: from adhesion molecules to signalling regulators.";
 RL Nat. Rev. Mol. Cell Biol. 4:33-45(2003).
 RN [16]
 RP PHOSPHORYLATION OF SER-706.
 RX MEDLINE=98248445; PubMed=9580567;
 RA Peck D., Isaacke C.M.;
 RT "Hyaluronan-dependent cell migration can be blocked by a CD44
 RT cytoplasmic domain peptide containing a phosphoserine at position
 RT 325.";
 RL J. Cell Sci. 111:1595-1601(1998).
 RN [17]
 RP PHOSPHORYLATION OF SER-672.
 RX MEDLINE=22038351; PubMed=12032545;
 RA Legg J.W., Lewis C.A., Parsons M., Ng T., Isaacke C.M.;
 RT "A novel PKC-regulated mechanism controls CD44 ezrin association and
 RT directional cell motility.";
 RL Nat. Cell Biol. 4:399-407(2002).
 RN [18]
 RP GLYCOSYLATION AND PROCESSING.
 RX MEDLINE=22766015; PubMed=12883358;
 RA Bartolozzi A.;
 RT "CD44 adhesive function spontaneous and PMA-inducible CD44 cleavage
 RT are regulated at post-translational level in cells of melanocytic
 RT lineage.";
 RL Melanoma Res. 13:325-337(2003).
 RN [19]
 RP VARIANT BLOOD GROUP INDIAN PRO-46.
 RX MEDLINE=96215152; PubMed=8636151;
 RA Telen M.J., Udani M., Washington M.K., Levesque M.C., Lloyd E.,
 RA Rao N.;
 RT "A blood group-related polymorphism of CD44 abolishes a hyaluronan-
 RT binding consensus sequence without preventing hyaluronan binding.";
 RL J. Biol. Chem. 271:7147-7153(1996).
 CC
 CC -1- FUNCTION: Receptor for hyalurononic acid (HA). Mediates cell-cell
 CC and cell-matrix interactions through its affinity for HA, and
 CC possibly also through its affinity for other ligands such as
 CC osteopontin, collagen, and matrix metalloproteinases (MMPs).
 CC Adhesion with HA plays an important role in cell migration, tumor
 CC growth and progression. Also involved in lymphocyte activation,
 CC rectification and homing, and in hematopoiesis. Altered expression
 CC or dysfunction causes numerous pathogenic phenotypes. Great
 CC protein heterogeneity due to numerous alternative splicing and
 CC post-translational modification events.
 CC
 CC -1- SUBUNIT: Interacts with HA, as well as other glycosaminoglycans,
 CC collagen, laminin, and fibronectin via its N-terminal segment.
 CC Interacts with ANK, the ERM proteins (VIL2, RDX and
 CC MSN), and NF2 via its C-terminal segment.
 CC
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist. Additional isoforms
 CC are produced by alternative splicing of 10 out of 19 exons
 CC within the extracellular domain. Additional diversity is

Query Match 94.7%; Score 2468.5; DB 1; Length 742;
 Best Local Similarity 66.3%; Pred. No. 2,5e-146;
 Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

QY 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTAADLCKAFNSTL 60
 DB 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTAADLCKAFNSTL 60
 QY 61 PTMAQMEKALSIGFETCRYGFLGHVAVIRIHPNSICANNNGVYLNTSQTYYCFN 120
 DB 61 PTMAQMEKALSIGFETCRYGFLGHVAVIRIHPNSICANNNGVYLNTSQTYYCFN 120
 QY 121 ASAPREDCTSVLDLBNADGPITITVNRDGRYOKGEYRTNEDIPSNPTDDVSS 180
 DB 121 ASAPREDCTSVLDLBNADGPITITVNRDGRYOKGEYRTNEDIPSNPTDDVSS 180
 QY 181 GSSESRSTSGGTYFTTSTVHPIDPESDPMITDSTRIPAT----- 222
 DB 181 GSSESRSTSGGTYFTTSTVHPIDPESDPMITDSTRIPAT----- 222
 QY 223 ----- 222
 DB 223 ----- 222
 QY 241 TMDWESWFLPSSKXHLHTTOMAGTSNTISAGMEPNEBENEDRHLSSGSGIDDD 300
 DB 241 TMDWESWFLPSSKXHLHTTOMAGTSNTISAGMEPNEBENEDRHLSSGSGIDDD 300
 QY 223 ----- 222
 DB 301 EDFISSTITTPPAFHTKONQDWTQWNSHNPVLLQTTMTMDVDNKGTAAYGNWN 360
 QY 223 ----- 222
 DB 361 PEAPHPLIHHEHEEETPHSTTIOATPSSITTEATQKQKQFNGRHEGRQTPREDS 420
 QY 223 -----NMDSHSTT 231
 DB 421 HSTTGTAASAATSHPMQGRTPSPEDSWTDFPNFISHPMRGHQAGRMMDSSHSTT 480
 QY 232 LQFTANPNGLVLDLRTGRLPLMTQOONSQSFSTHEDLEDKKHPTSTLTSSNRDV 291
 DB 481 LQFTANPNGLVLDLRTGRLPLMTQOONSQSFSTHEDLEDKKHPTSTLTSSNRDV 540
 QY 292 TGGRRDPNHSSEGSTTLLEGYTSHPYTKESRTFIPYTSKTSFGVTAAYVDSNSNVNR 351
 DB 541 TGGRRDPNHSSEGSTTLLEGYTSHPYTKESRTFIPYTSKTSFGVTAAYVDSNSNVNR 600
 QY 352 SLSGDDDTFRPSGSHTHGSESDGSHSGSEGGANTTSGPRTQPPIPWLLIILASLAL 411
 DB 601 SLSGDDDTFRPSGSHTHGSESDGSHSGSEGGANTTSGPRTQPPIPWLLIILASLAL 660
 QY 412 ALLIACIANSRRRGQKKLVINSNGAVEDRKPSSGLNGEASQSEWHLVNNKESST 471
 DB 661 ALLIACIANSRRRGQKKLVINSNGAVEDRKPSSGLNGEASQSEWHLVNNKESST 720
 QY 472 PDQFTADERTANQNTDMKIGV 493
 DB 721 PDQFTADERTANQNTDMKIGV 742

RESULT 2
 CD44_PAPHA STANDARD; PRT; 362 AA.
 AC P14745;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-MAR-1990 (Rel. 14, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE CD44 antigen precursor (Phagocytic glycoprotein I) (PCG-1) (HUTCH-1)
 DE (extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte
 DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor).
 GN CD44.
 OS Papio hamadryas (Hamadryas baboon).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Papio.

NCBI_TaxID=9557;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 66-74.
 RX MEDLINE=69282830; PubMed=2471974;
 RA Idetia R.L., Carter W.G., Nottenburg C., Wayner E.A.,
 RA Gallatin W.M., St John T.J.
 RT "Isolation and DNA sequence of a cDNA clone encoding a lymphocyte
 RT adhesion receptor for high endothelium."
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4659-4663 (1989).
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
 CC mucosal high endothelial venule and to types I and VI collagen.
 CC Probably involved in matrix adhesion, lymphocyte activation and
 CC lymph node homing.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: Extensively modified including N- and O-linked glycosylation,
 CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,
 CC of phosphate to cytoplasmic domain serine residues.
 CC -1- SIMILARITY: Contains 1 link domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 CC EMBL; M2452; AAA35385.1; -.
 DR HSSP; P98066; ITSG.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRODOM; PD000918; Link; 1.
 DR SMART; SM00445; LINK; 1.
 DR PROSITE; PS01241; LINK; 1.
 KW Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
 KW Proteoglycan; Signal; Alternative splicing;
 KM Pyroglutamate carboxylic acid.
 FT SIGNAL 1 20
 FT CHAIN 21 362
 FT DOMAIN 21 269
 FT TRANSMEM 270 290
 FT DOMAIN 291 362
 FT DOMAIN 48 119
 FT DOMAIN 150 158
 FT DOMAIN 225 269
 FT MOD_RES 21 21
 FT DISULFID 53 118
 FT DISULFID 77 97
 FT MOD_RES 292 292
 FT MOD_RES 326 326
 FT CARBOHYD 25 25
 FT CARBOHYD 57 57
 FT CARBOHYD 100 100
 FT CARBOHYD 110 110
 FT CARBOHYD 120 120
 FT CARBOHYD 256 256
 FT CONFLICT 67 67
 SQ SEQUENCE 362 AA; 39378 MW; 578BCE/C3D52EFF CRC64;

Query Match 65.1%; Score 1698.5; DB 1; Length 362;
 Best Local Similarity 68.2%; Pred. No. 7.7e-99;
 Matches 337; Conservative 7; Mismatches 17; Indels 133; Gaps 2;

QY 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTAADLCKAFNSTL 60
 DB 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTAADLCKAFNSTL 60
 QY 61 PTMAQMEKALSIGFETCRYGFLGHVAVIRIHPNSICANNNGVYLNTSQTYYCFN 120

DB 61 PTMAQWELKLSIGETRCRGYFISGHVYIPRIHENSICANNVTGVIITLTSNTSGYDTYCN 120
QY 121 ASAPPEDCTSYVDLPNAPGPTTITVNDGTRVYQKGRYRNPEDIPSNPTDDVS 180
DB 121 ASAPGEDCTSYVDLPNAPGPTTITVNDGTRVYQKGRYRNPEDIPSNPTDDVS 180
QY 181 GSSSERSTSGYFYFT-FSTVHPEDPESPWITDSTRIPATNMDSSHSTLQPTANEN 239
DB 181 GSSSERSTLGGYFYFVHFSTSPIDPDEGPTWITDSTRIPATNMDSSHSTLQPTANEN 224
QY 240 TGLVEDDRTGPTSLMTTQGSNSQSFSTHSGLEDKDPHTSTLTSSNNDVTGGRDN 239
DB 225 ----- 224
QY 300 HSEGSTLLLEGYTSHPHYTESRTFIPVTSKAGTSFGVAVTVGDNSNVNRSLSGDQDT 359
DB 225 -----DQGA 228
QY 360 FHPSGSHHTTGSSDSHSGSGEGANTSGPRTQPIPEWIIILASLALALILANCI 419
DB 229 FDPGSGHTTGSSESASHSGSREGANTSGPLTPQPIPEWIIILASLALALILANCI 288
QY 420 AVNSRRRCQCKKIVINSNGAVEDRKPSSGLNGBASQSMVHLVKNSESETPDQFMTAD 479
DB 289 AVNSRRRCQCKKIVINSNGAVEDRKPSSGLNGBASQSMVHLVKNSESETPDQFMTAD 348
QY 480 ETRNLQNVDMKIGV 493
DB 349 ETRNLQNVDMKIGV 362

RESULT 3
CD44 MOUSE STANDARD; FRT; 778 AA.
ID CD44 MOUSE
AC P15379; Q05732; Q61395; Q62060; Q62061; Q62062; Q62063; Q62408;
AC Q62409; Q64296; Q99114; Q991X8;
AC 01-APR-1990 (Rel. 14, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CD44 antigen precursor (phagocytic glycoprotein I) (Pgp-1) (HUTCH-I)
DE (extracellular matrix receptor-III) (ECCR-III) (GP90 lymphocyte
DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)
DE (IX-24).
GN CD44.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS 4; 6; 7 AND 12).
RC STRAIN=DMB/2; TISSUE=Lung;
RC MEDLINE=93107170; PubMed=1469058;
RA He O., Lesley J., Hyman R., Ishihara K., Kinsade P.W.;
RA "Molecular isoforms of murine CD44 and evidence that the membrane
RA proximal domain is not critical for hyaluronate recognition.";
RT J. Cell Biol. 119:1711-1719(1992).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM 13).
RX MEDLINE=90038499; PubMed=2681416;
RA Zhou D.F.H., Ding J.F., Pickler L.J., Bargatze R.F., Butcher E.C.,
RA Goeddel D.V.;
RA "Molecular cloning and expression of Pgp-1. The mouse homolog of the
RA human H-CAM (Hermes) lymphocyte homing receptor.";
RL U. Immunol. 143:3390-3395(1989).
RN [3]
RN SEQUENCE FROM N.A. (ISOFORM 13).
RX MEDLINE=90046829; PubMed=2682651;
RA Notenberg C., Rees G., St John T.;
RA "Isolation of mouse CD44 cDNA: structural features are distinct from
RA the primate CDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8521-8525(1989).
RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20318634; PubMed=10859330;
RA Witting B.M., Johansson B., Zoeller M., Schwaerzler C., Guentert U.;
RT "Abrogation of experimental colitis correlates with increased
RT apoptosis in mice deficient for CD44 variant exon 7 (CD44v7).";
RL J. Exp. Med. 191:2053-2064(2000).
RN [5]
RN SEQUENCE FROM N.A. (ISOFORM 13).
RX MEDLINE=22388257; PubMed=12477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Stensberg R.D., Collins F.S., Wagner L., Sherman C.M., Schuller G.D.,
RA Alechul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant I.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RN SEQUENCE FROM N.A. (ISOFORM 13).
RP STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nixardo I., Oseko N., Saito R., Suzuki H., Yamana K., Kiyosawa H.,
RA Yagi K., Tomaru I., Hasegawa Y., Nogami A., Schombach C., Gojodori T.,
RA Balakrishna R., Hill D.P., Bult C., Hume D.A., Quackenbush U.,
RA Schiml L.M., Kanapin A., Matveeva H., Batalov S., Beisel K.W.,
RA Blake J.A., Brad D., Brusic V., Chotcha C., Corbani L.B., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Prazner K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Gough J.,
RA Grimmond S., Guerinich S., Hirokawa N., Jackson I.T., Jarvis E.D.,
RA Kani A., Kawaji H., Kawasawa Y., Kedziarski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lehner B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Petrea G., Pesole G.,
RA Pavlovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Raveisi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Warande Y., Wells C.,
RA Wilting L.G., Wynshaw-Boris A., Yangisawa M., Yang I., Yang L.,
RA Yuan Z., Zavoian M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imocani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [7]
RN SEQUENCE OF 8-778 FROM N.A. (ISOFORM 13).
RP MEDLINE=90094420; PubMed=2403559;
RX Wolfe E.U., Gause W.C., Peitrey C.M., Holland S.M., Steinberg A.D.,
RA August J.T.;
RA "The cDNA sequence of mouse Pgp-1 and homology to human CD44 cell
RT surface antigen and proteoglycan core/link proteins.";
RL J. Biol. Chem. 265:341-347(1990).
RN [8]
RN SEQUENCE OF 224-637 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6; 7 AND 8).
RC STRAIN=GR;

RX MEDLINE=93219085; PubMed=8464707;
 RA Toelg C., Hofmann M., Herrlich P., Ponta H.;
 RT "Splicing choice from ten variant exons establishes CD44
 RT variability.";
 RL Nucleic Acids Res. 21:1225-1229(1993).
 [9]
 RP SEQUENCE OF 224-637 FROM N.A. (ISOFORM 9).
 RC STRAIN=BALB/c;
 RA MEDLINE=93286043; PubMed=8509359;
 RA Screaton G.R., Bell M.V., Bell J.I., Jackson D.G.;
 RT "The identification of a new alternative exon with highly restricted
 RT tissue expression in transcripts encoding the mouse Pgp-1 (CD44)
 RT homing receptor. Comparison of all 10 variable exons between mouse,
 RT human, and rat.";
 RL J. Biol. Chem. 268:12235-12238(1993).
 [10]
 RN PARTIAL SEQUENCE FROM N.A. (ISOFORMS 10 AND 11).
 RP STRAIN=Swiss Webster;
 RC MEDLINE=96355396; PubMed=8702806;
 RA Yu Q., Toole B.P.;
 RT "A new alternatively spliced exon between v9 and v10 provides a
 RT molecular basis for synthesis of soluble CD44.";
 RL J. Biol. Chem. 271:20603-20607(1996).
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
 CC mucosally high endothelial venule and to types I and VI collagen.
 CC Probably involved in matrix adhesion, lymphocyte activation and
 CC lymph node homing.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=13;
 CC Name=1;
 CC IsoId=P15379-14; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P15379-7; Sequence=VSP_007329;
 CC Name=3;
 CC IsoId=P15379-8; Sequence=VSP_007330;
 CC Name=4; Synonyms=M2;
 CC IsoId=P15379-4; Sequence=VSP_007331;
 CC Name=5;
 CC IsoId=P15379-9; Sequence=VSP_007332;
 CC Name=6; Synonyms=M3;
 CC IsoId=P15379-5; Sequence=VSP_005326;
 CC Name=7; Synonyms=M4;
 CC IsoId=P15379-6; Sequence=VSP_005327;
 CC Name=8;
 CC IsoId=P15379-10; Sequence=VSP_007330; VSP_007334;
 CC Name=9;
 CC IsoId=P15379-11; Sequence=VSP_007332; VSP_007335;
 CC Name=10;
 CC IsoId=P15379-12; Sequence=VSP_007336; VSP_007337;
 CC Name=11;
 CC IsoId=P15379-13; Sequence=VSP_007338; VSP_007339;
 CC Name=12; Synonyms=M1;
 CC IsoId=P15379-3; Sequence=VSP_005328;
 CC Name=13; Synonyms=M0;
 CC IsoId=P15379-2; Sequence=VSP_005329;
 CC -1- PTM: N-glycosylated (by similarity);
 CC -1- PTM: O-glycosylated; contains chondroitin sulfate glycans which
 CC can be more or less sulfated (By similarity).
 CC -1- PTM: Phosphorylated; activation of PKC results in the
 CC dephosphorylation of Ser-742 (constitutive phosphorylation site),
 CC and the phosphorylation of Ser-708 (By similarity).
 CC -1- POLYMORPHISM: Two allelic forms of this glycoprotein, PGP-1.1 and
 CC PGP-1.2, have been reported. The expressed product is PGP-1.1 (ly-
 CC 24.1).
 CC -1- SIMILARITY: Contains 1 link domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC DR EMBL; X66084; CAA46883.1; -
 CC DR EMBL; X66083; CAA46882.1; -
 CC DR EMBL; X66082; CAA46881.1; -
 CC DR EMBL; X66081; CAA46880.1; -
 CC DR EMBL; M30655; AAA39922.1; -
 CC DR EMBL; M27129; AAA37406.1; -
 CC DR EMBL; M27130; AAA37407.1; -
 CC DR EMBL; AJ251594; CAB61868.1; -
 CC DR EMBL; BC005676; AAH05676.1; -
 CC DR EMBL; AK045226; BAC32269.1; -
 CC DR EMBL; J05163; AAA39923.1; -
 CC DR EMBL; X69724; CAA49380.1; -
 CC DR EMBL; U13611; AAA37445.1; -
 Query Match 63.7%; Score 1662.5; DB 1; Length 778;
 Best Local Similarity 47.4%; Pred. No. 3; 6e-96;
 Matches 370; Conservative 39; Mismatches 82; Indels 289; Gaps 10;
 CC -----
 CC 1 MDKFWHAAWGLCVPLSLA--QIDNITCRFAGVFHYKXGRYSISRTTEADLCXAFNS 58
 CC 1 MDKFWHAAWGLCVPLSLA--QIDNITCRFAGVFHYKXGRYSISRTTEADLCXAFNS 60
 CC 59 TLPTMAQWEKALSIGFETCRGYFEGHVIPIRIHPSICAAANTGVYL-TSNTSQYDTY 117
 CC 61 TLPTMDQWKLALSKGFETCRGYFEGHVIPIRIHPSICAAANTGVYLVTNNTSHYDTY 120
 CC 118 CFNASAPPEEDCTSVTLNPAFDPPIITTVNPGCTGYVKGERTUPEDIPENPDD 177
 CC 121 CFNASAPPEEDCTSVTLNPAFDPPIITTVNPGCTGYVKGERTUPEDIPENPDD 180
 CC 178 VSSGSSSRSTSGCYLYFP-----STVH----- 202
 CC 181 VSSGSLTEK-STPGYLITHTLPTPEQTPGDDSFIRSTLATIATVSKSHAAQKON 239
 CC 203 ----- 202
 CC 240 NWIWSFGNSQSTQTOEPTTSATVLTMTPTPKQOEAQNFWSLFPQSESKSHHTT 299
 CC 203 ----- 202
 CC 300 TKMPOTESNTPTQWEPNEBNEDETDYPPSPSGSGIDDDDEFISITATTPRVGARTEDN 359
 CC 203 ----- 202
 CC 360 QDWTQWKENHNPVLLQTTTRMADIDRISTSAHGENWTPPEPPFNNHEXODEETPHA 419
 CC 203 -----PIPEDS----- 209
 CC 420 TSTTPNSTAAEAATQOETWPFONGQGNPPPIPSDSHVITAGTTASANNHNSCRITTSQ 479
 CC 210 ---PWITDSTDR-----DATMDSHSTTLTOPANPNTGLVEDLDRTPPLAMTT 256
 CC 480 EDVGSW-TDFPDPISHPMQOQHQSXTDSSHSSTLPTAPFNHVLVEDLNRTPLPVTT 538
 CC 257 QQNSQSFSTSHGEGLEEDKHPTSTLTSNRANDVTGGRDPNHSBSSTTLLEGYTSHY 316
 CC 539 PQSHQNSSTLHGEPEEDENFTTSLIPSTKSKAKARQGSIPDTTTSVEGYTQYF 598
 CC 317 HTKESRPIPTYSAKTGSFGVTAIVTG-DSNSNVNRLSGDQTFPHSGSH--TTGSE 373
 CC 599 DTMEGTLFPTPAKTEVFGETEVLTLTDSNVNVDGLPGDRDSKDSRGSRTVTGSE 658
 CC 374 SDGSHSGQSGAANTTSGPIRTPIPEMLIILASLALALIALAVCIANSRRRCQKKL 433
 CC 659 LAGHSNAQDSGVTTTSGEMRPPIPERILLALALALALAVCIANSRRRCQKKL 718
 CC 434 VINSNGAVEDRPSKINGEASKQSENVHLVKNSSSTPDQFMATDETRNLQNVDMKIGV 493
 CC 719 VINGNGVIEDRPSKINGEASKQSENVHLVKNSSSTPDQFMATDETRNLQNVDMKIGV 778

ID	CD44_MESAU	STANDARD	PRF	431 AA.
AC	CD44_MESAU			
AD	060522; 060523;			
AE	01-NOV-1997 (Rel. 35, Created)			
AF	01-NOV-1997 (Rel. 35, Last sequence update)			
AG	15-MAR-2004 (Rel. 43, Last annotation update)			
AH	CD44 antigen precursor (Phagocytic glycoprotein I) (GPG-I) (HUTCH-I)			
AI	(Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte			
AJ	homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)			
AK	(Heparan sulfate proteoglycan) (HAMI antigen).			
AL	CD44			
AM	Mesocricetus auratus (Golden hamster).			
AN	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AO	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Circetinae;			
AP	Mesocricetus.			
AQ	NCBI_TaxID=10036;			
AR	[1]			
AS	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).			
AT	STRAIN-LING, TISSUE=Alveolar macrophage;			
AV	Paulauskis J.D., Kobzik L., Gerard C., Katler M., Godleski J.J.;			
AW	Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.			
AX	-1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to			
AY	mucosal high endothelial venule and to types I and VI collagen.			
AZ	Probably involved in matrix adhesion, lymphocyte activation and			
BA	lymph node homing.			
BB	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
BC	-1- ALTERNATIVE PRODUCTS:			
BD	Event=Alternative splicing; Named isoforms=2;			
BE	Comment=Additional isoforms seem to exist;			
BF	Name=1;			
BG	Isoid=060522-1; Sequence=Displayed;			
BH	Name=2;			
BI	Isoid=060522-2; Sequence=VSP 005322;			
BJ	-1- PTM: N-glycosylated (By similarity).			
BK	-1- PTM: O-glycosylated. Contains chondroitin sulfate glycans which			
BL	can be more or less sulfated (By similarity).			
BM	-1- PTM: Phosphorylated; activation of SRC results in the			
BN	dephosphorylation of Ser-395 (constitutive phosphorylation site),			
BO	and the phosphorylation of Ser-361 (By similarity).			
BP	-1- SIMILARITY: Contains 1 link domain.			
BQ	-----			
BR	This SWISS-PROT entry is copyright. It is produced through a collaboration			
BS	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
BT	the European Bioinformatics Institute. There are no restrictions on its			
BU	use by non-profit institutions as long as its content is in no way			
BV	modified and this statement is not removed. Usage by and for commercial			
BW	entities requires a license agreement (See http://www.isb-sdb.ch/announce			
BX	or send an email to license@isb-sdb.ch).			
BY	-----			
CA	EMBL; U10880; AAC13767.1; -			
CB	EMBL; U10881; AAA19316.1; -			
CC	HSSP; P98066; ITSG.			
CD	InterPro; IPR001231; CD44_antigen.			
CE	InterPro; IPR000538; Link.			
CF	Pfam; PF00193; XLink; 1.			
CG	PRINTS; PR00658; CD44.			
CH	PRINTS; PR01265; LINKMODULE.			
CI	ProDom; PD000918; Link; 1.			
CJ	SMART; SM00445; Link; 1.			
CK	PROSITE; PS01241; Link; 1.			
CL	Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;			
CM	Proteoglycan; Signal; Alternative splicing;			
CN	Pyroglutamate; carboxylic acid.			
CO	FT SIGNAL	1	22	
CP	FT CHAIN	23	431	BY SIMILARITY.
CQ	FT DOMAIN	23	338	CD44 ANTIGEN.
CR	FT TRANSMEM	339	359	EXTRACELLULAR (POTENTIAL).
CS	FT DOMAIN	360	431	POTENTIAL.
CT	FT DOMAIN	50	121	CYTOPLASMIC (POTENTIAL).
CU	FT DOMAIN	152	160	LINK.
CV	FT DOMAIN	226	338	ARG/LYS-RICH (BASIC).
CA	MOD RES	23		STEM.
CB				PYROLIDONE CARBOXYLIC ACID (BY

FT	DISULFID	55	120	BY SIMILARITY.
FT	DISULFID	79	99	BY SIMILARITY.
FT	MOD_RES	361	361	PHOSPHORYLATION (BY PKC) (PARTIAL) (BY SIMILARITY)
FT	MOD_RES	395	395	PHOSPHORYLATION (PARTIAL) (BY SIMILARITY)
FT	CARBOHYD	27	27	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	59	59	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	102	102	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	112	112	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	122	122	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	174	174	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	256	256	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	325	325	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	VARSPLIC	222	291	TRSGCKGRRGGGLPKCATS . . .) (POTENTIAL)
FT				TPATGVGGEIEVVAEDSNVDSLSLG -> R (1n isoform 2)
FT				/Ftld-VSP 005322.
SEQ	SEQUENCE	431 AA;	46807 MM;	4300262BDC6BEA6A CRC64;
Query Match	62.5%;	Score 1631;	DB 1;	Length 431;
Best Local Similarity	67.5%;	Pred. No. 1.5e-94;		
Matches 337;	Conservative	28;	Mismatches 60;	Indels 74;
				Gaps 9
QY	1	MDKFWMAHAGLCVPLSLIA--QIDLNTCCFAAGVFHNEKNGRYSISRTEADLCKAENS	58	
DB	1	MDKFWMAHAGLCVPLSLIAQQQIDLNTCCFAAGVFHNEKNGRYSISRTEADLCQAENS	60	
QY	59	TLPTMAQWEKALSTGFETCRGFIIEGHVIVIRIHNSICANNNGVYLLTNTSOYDNC	118	
DB	61	TLPTMDQWVMAISKFEFCRGFIIEGHVIVIRIOPNALCAANHGVYLLTNTSHYDNC	120	
QY	119	FNASAPREDCSVYDLDPNAPFDGPTITIVNRDGRYVQKGEYRTPNEDIYFSPNTPDDV	178	
DB	121	FNASAPLEDDCSVYDLDPNAPFDGPTITIVNRDGRYVQKGEYRTHQEDIDPASNTPDDV	180	
QY	179	SGSSSESSSTSGGVIETFTF--STYHPRIDEDSPMTDSDTRIPATMSSSHATLQPTAN	237	
DB	181	SGSSSEK-STSGGVFHTVYPLTSHDQDDPFTIGST--	218	
QY	238	PNTGLVEDLDRGTPLSMTTQOONSQGSFSTSHGLEDXDHPTTSLTSSNRDVTGRRD	297	
DB	219	-----MATRSGG-----KD-----GRGG	232	
QY	298	PWHSRGTTLLEGYTSYHPHTRKESPTFIPYSAKTSQGVAVTVG--DQSNVNRSLGSD	356	
DB	233	GGLPDQATISLGGYTHTRPETWENQTLIPVPAKTVGGEIEVVAEDSNFVDSLPED	292	
QY	357	QD-TTHPSGSGHT-THGSDSDSHSGSQEGGANTTSGPIRTPQIPDEWLIILASLALAI	414	
DB	293	QDSMDPGRNSTLTVDDGSKLTGHSNGDQSDGANTTSRGRKQIPDEWLIILASLALAI	352	
QY	415	LAVCAVNASRRRCGGKKGLVINSNGAVEDRPSGLNGEASQSEWHLVNRKESSETPDQ	474	
DB	353	LAVCAVNASRRRCGGKKGLVINSNGAVEDRPSGLNGEASQSEWHLVNRKESSETPDQ	412	
QY	475	FMTADETRNLQNVDMKIGV	493	
DB	413	FMTADETRNLQNVDMKIGV	431	
RESULT 5	CD44_HORSE	STANDARD;	PRT;	359 AA.
ID	CD44_HORSE	005078;		
AC	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	CD44 antigen precursor (Phagocytic glycoprotein 1) (Pgp-1) (HUTCH-1)			
DE	(Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte			
DE	heming/adhesion receptor) (Hermes antigen) (Hyaluronate receptor).			
GN	CD44.			

OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=3170897; Pubmed=8436424;
 RA Tavernor A.S., Deverson E.V., Coadwell W.J., Lunn D.P., Zhang C.,
 RA Davis W., Butcher G.W.,
 RT "Molecular cloning of equine CD44 cDNA by a COS cell expression
 RT system";
 RL Immunogenetics 37:474-477(1993).
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
 CC mucosal high endothelial venule and to types I and VI collagen.
 CC Probably involved in matrix adhesion, lymphocyte activation and
 CC lymph node homing.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: Extensively modified including N- and O-linked glycosylation,
 CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,
 CC of phosphate to cytoplasmic domain serine residues.
 CC -1- SIMILARITY: Contains 1 link domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X66862; CAA47331.1; -;
 DR F01; 146245; S24240.
 DR HSP; P98066; 1TSG.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PRO0658; CD44.
 DR PRINTS; PRO1265; LINKMODULE.
 DR ProDom; PD000918; Link; 1.
 DR SMART; SM00445; LINK; 1.
 DR PROSITE; PS01241; LINK; 1.
 KW Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
 KW Proteoglycan; Signal; Alternative splicing;
 KW Pyrrolidone carboxylic acid.
 KW SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 1 359 CD44 ANTIGEN.
 FT DOMAIN 21 266 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 267 287 POTENTIAL.
 FT DOMAIN 288 359 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 48 119 LINK.
 FT DOMAIN 150 158 ARG/LYS-RICH (BASIC).
 FT DOMAIN 225 266 STEM.
 FT MOD_RES 21 21 PYRROLIDONE CARBOXYLIC ACID (BY
 FT SIMILARITY).
 FT DISULFID 53 118 BY SIMILARITY.
 FT DISULFID 77 97 BY SIMILARITY.
 FT MOD_RES 289 289 PHOSPHORYLATION (BY PKC) (PARTIAL) (BY
 FT SIMILARITY).
 FT MOD_RES 323 323 PHOSPHORYLATION (PARTIAL) (BY
 FT SIMILARITY).
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 359 AA; 38990 MW; BE20461C587AA34B CRC64;
 Query Match 59.3%; Score 1547; DB 1; Length 359;
 Best Local Similarity 63.0%; Pred. No. 2e-89; Indexes 138; Gaps 5;
 Matches 312; Conservative 20; Mismatches 25;

DB 1 MDKFWMAAMGLCVPLSLAQIDINTICRAGVHVHVKRNRYSISRTREADLCKAFNSTL 60
 QY 61 PTMAQMERKALSIGETGRYGFIEGHVVIPIHNSICAAANTGYIITSTSCDYDYG 120
 DB 61 PTMAQMERKALNIGETGRIGIEGHVVIPIHNSICAAANTGYIITSTSCDYDYG 120
 QY 121 ASAPPEEDCTSVTLPAFADPPIITITVNRDGTGYVOKGEYRTNVEDIYSPNPDVVSS 180
 DB 121 ASAPPEEDCTSVTLPAFADPPIITITVNRDGTGYVOKGEYRTNVEDIYSPNPDVVSS 180
 QY 181 GSSSERSTSGCY-IFVT-PSYVAPIDEDSPWITDSTRIPATNDSSTTLQPTAMP 238
 DB 181 GSSSER-STSGCYVIFHTHTPTTPTDQSSPWSDBPKPTPTT----- 223
 QY 239 MTGLVEDIDRGPISMTTQGSNSGSPFSTHGLEEDKHPTTSTLTSSNRDVTGRRDP 298
 DB 224 -----KORASGR--- 231
 QY 299 NHSEGSTTLLEGYTSYPTKESRTFIPVTSAKTSGVYAVTVGDSNVNBSLSDQD 358
 DB 232 ----- 231
 QY 359 TFHSGSHHTHSGESDGHSGOEGANTTGPRTTPOIPEWTLIASILALALILAVC 418
 DB 232 -----AQTHSGSETGSHSGOEGASTTSGIRRPQIPWTLIASILALALILAVC 284
 QY 419 IAVNSRRRCGKKLVINSNGAVEDRKPGLNGEASKSQEMVTLVKNESSETPDQMTA 478
 DB 285 IAVNSRRRCGKKLVINSNGAVEDRKPGLNGEASKSQEMVTLVKNESSETPDQMTA 344
 QY 479 DETRLQNVDMKIEV 493
 DB 345 DETRLQNVDMKIEV 359
 RESULT 6
 ID CD44_CANPA STANDARD; PRT; 351 AA.
 GN CD44_CANPA
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Beagle; TISSUE=Thymus;
 RC MEDLINE=94250687; Pubmed=7514890;
 RA Milde K.F., Alejandro R., Mintz D.H., Pastori R.L.,
 RT "Molecular cloning of the canine CD44 antigen cDNA".
 RT Biochim. Biophys. Acta 1218:112-114(1994).
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
 CC mucosal high endothelial venule and to types I and VI collagen.
 CC Probably involved in matrix adhesion, lymphocyte activation and
 CC lymph node homing.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Lymph nodes.
 CC -1- PTM: Extensively modified including N- and O-linked glycosylation,
 CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,
 CC of phosphate to cytoplasmic domain serine residues (By
 CC similarity).
 CC -1- SIMILARITY: Contains 1 link domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

FT DOMAIN 21 273 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 274 294 POTENTIAL.
FT DOMAIN 295 366 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 448 119 LINK.
FT DOMAIN 150 158 ARG/LYS-RICH (BASIC).
FT DOMAIN 229 273 STEM.
FT MOD_RES 21 21 PYROLIDONE CARBOXYLIC ACID (BY
FT SIMILARITY).
FT DISULFID 53 118 BY SIMILARITY.
FT DISULFID 77 97 BY SIMILARITY.
FT MOD_RES 296 296 PHOSPHORYLATION (BY PKC) (PARTIAL) (BY
FT SIMILARITY).
FT MOD_RES 330 330 PHOSPHORYLATION (PARTIAL) (BY
FT SIMILARITY).
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 366 AA; 40001 MW; 438A5A1E631E02B4 CRC64;

```

```

Query Match 56.3%; Score 1469.5; DB 1; Length 366;
Best Local Similarity 61.0%; Pred. No. 1.4e-84;
Matches 305; Conservative 22; Mismatches 32; Indels 14; Gaps 8;

```

```

QY 1 MDKFWHAAMGICLVEPLSLAIDILNITCRFAGVFEYKNGRYSISRTAADICKAFNSTL 60
DB 1 MDTFWRAAMGICLVEPLSLAIDILNITCRFAGVFEYKNGRYSISRTAADICKAFNSTL 60
QY 61 PTAAOMKALISGFECRCYGFIEGHVIRIHPNISCANNTGVYILNITSQDYTCGN 120
DB 61 PTAAOMKALISGFECRCYGFIEGHVIRIHPNISCANNTGVYILNITSQDYTCGN 120
QY 121 ASAPREDCTSVTLDPNADGPITITIVNRDGRVYQKGYRTNEDI-----YPSNPTD 175
DB 121 ASAPREDCTSVTLDPNADGPITITIVNRDGRVYQKGYRTNEDI-----YPSNPTD 175
QY 121 ASAPREDCTSVTLDPNADGPITITIVNRDGRVYQKGYRTNEDI-----YPSNPTD 180
DB 121 ASAPREDCTSVTLDPNADGPITITIVNRDGRVYQKGYRTNEDI-----YPSNPTD 180
QY 176 DVVSSGSSSRSTSGY-IEYV-FETVHPIDPDSPTITDTRIPANMDSSTITIQ 233
DB 176 DVVSSGSSSRSTSGY-IEYV-FETVHPIDPDSPTITDTRIPANMDSSTITIQ 233
QY 181 DESSGSSPSEK-STSGYSLFPHLPTVPSRPR-RPW-----SQREKNT----- 223
DB 181 DESSGSSPSEK-STSGYSLFPHLPTVPSRPR-RPW-----SQREKNT----- 223
QY 234 PTANPTVGLVEDLDRTPGSLMTQOSNQSFSYSHGLEEDKHPTSTLTSSNNDVYG 293
DB 234 PTANPTVGLVEDLDRTPGSLMTQOSNQSFSYSHGLEEDKHPTSTLTSSNNDVYG 293
QY 224 -----SDTRDYGSSHD----- 234
DB 224 -----SDTRDYGSSHD----- 234
QY 294 GRADPNHSGSTLLLEGYTSHPYTKESRTFIVISAKTGSFGVAVTVGDSNSVNSL 353
DB 294 GRADPNHSGSTLLLEGYTSHPYTKESRTFIVISAKTGSFGVAVTVGDSNSVNSL 353
QY 235 ----- 234
DB 235 ----- 234
QY 354 SGDDTFHPSGSHHTGSEDSHSGOEGANTTSGPIRTPOIPEWLIILASLALAL 413
DB 354 SGDDTFHPSGSHHTGSEDSHSGOEGANTTSGPIRTPOIPEWLIILASLALAL 413
QY 235 -----PSGRSYTHASASGSSGSEHGANTTSGPMKQPIDEWLIILASLALAL 286
DB 235 -----PSGRSYTHASASGSSGSEHGANTTSGPMKQPIDEWLIILASLALAL 286
QY 414 ILAVCIANRRRCGQKKCLVINSNGAVEDRKPSGLNGEASKSGQMVHLVKNSESPTD 473
DB 414 ILAVCIANRRRCGQKKCLVINSNGAVEDRKPSGLNGEASKSGQMVHLVKNSESPTD 473
QY 287 ILAVCIANRRRCGQKKCLVINSNGAVEDRKPSGLNGEASKSGQMVHLVKNSESPTD 346
DB 287 ILAVCIANRRRCGQKKCLVINSNGAVEDRKPSGLNGEASKSGQMVHLVKNSESPTD 346
QY 474 QPMTADETRLQNVDMKIGV 493
DB 474 QPMTADETRLQNVDMKIGV 493
QY 347 QPMTADETRLQNVDMKIGV 366
DB 347 QPMTADETRLQNVDMKIGV 366

```

```

DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)
DE (LY-24).
OS CD44.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=BDIX; TISSUE=Pancreas;
RX MEDLINE=9119152; Pubmed=1707342;
RA Guenther U., Hofmann M., Rudy W., Reber S., Zoeller M.,
RA Hausmann I., Natzku S., Wenzel A., Ponta H., Herrlich P.,
RA "A new variant of glycoprotein CD44 confers metastatic potential to
RA rat carcinoma cells."
RT Cell 65:13-24 (1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Stevens J.W., Midura R.J.;
RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
RL mucosal high endothelial venule and to types I and VI collagen.
RL Probably involved in matrix adhesion, lymphocyte activation and
RL lymph node homing.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=2; Synonyms=Long, Meta-1;
CC IsoId=P26051-1; Sequence=Displayed;
CC Name=1; Synonyms=Short;
CC IsoId=P26051-2; Sequence=VSP_005330;
CC -1- PTM: N-glycosylated (By similarity).
CC -1- PTM: O-glycosylated; contains chondroitin sulfate glycans which
CC can be more or less sulfated (By similarity).
CC -1- PTM: Phosphorylated; activation of PKC results in the
CC dephosphorylation of Ser-467 (constitutive phosphorylation site),
CC and the phosphorylation of Ser-433 (By similarity).
CC -1- SIMILARITY: Contains 1 link domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M61875; AAAS3532.1; -
CC EMBL; M61874; AAAS3534.1; -
CC EMBL; U52179; AAA97915.1; -
CC EMBL; U46957; AAA92920.1; -
CC PIR; B38745; B38745.
CC HSPD; P98066; ITSG.
CC InterPro; IPR001231; CD44_antigen.
CC InterPro; IPR00538; Link.
CC Pfam; PF00193; Xlink; 1.
CC PRINTS; PR00658; CD44.
CC PRINTS; PR01265; LINKMODULE.
CC Prodom; PD000918; Link; 1.
CC SMART; SM00445; LINK; 1.
CC PROSITE; PS01241; LINK; 1.
CC Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
CC Proteoglycan; Signal; Alternative splicing;
CC Pyroglutamate; Pyroglutamate;
CC SIGNAL 1 21
CC CHAIN 22 503
CC DOMAIN 22 410
CC TRANSMEM 411 431
CC DOMAIN 432 503
CC DOMAIN 51 123
CC DOMAIN 154 162
CC DOMAIN 228 410
CC STEM.

```

```

RESULT 8
CD44_RAT STANDARD; PRT; 503 AA.
AC P26051; 099021;
AC 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CD44 antigen precursor (Phagocytic glycoprotein I) (Pgp-I) (HUTCH-I)
DE (Extracellular matrix receptor-III) (ECMR-III) (Gp90 lymphocyte

```

[illegible]

DE	CD44 antigen precursor (Phagocyte glycoprotein I) (PGP-1) (HMTCH-1)
DE	(Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte
DE	homing/adhesion receptor) (Hermes antigen) (hyaluronate receptor).
GN	CD44.
OS	Cricetulus griseus (Chinese hamster).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
CC	Cricetus.
CC	NCBI_TaxId=10029;
RN	[1]
RX	SEQUENCE FROM N.A.
RX	MEDLINE=90304889; PubMed=1694723,
RA	Arafat A., Stamenkovic I., Melnick M., Underhill C.B., Seed B.;
RT	"CD44 is the principal cell surface receptor for hyaluronate";
RL	Cell 61:1303-1313(1990).
CC	-1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
CC	mucosal high endothelial venule and to types I and VI collagen.
CC	Probably involved in matrix adhesion, lymphocyte activation and
CC	lymph node homing.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- INDUCTION: By EBV.
CC	-1- PTM: Extensively modified including N- and O-linked glycosylation,
CC	addition of the glycosaminoglycan chondroitin sulfate, of sulfate,
CC	of phosphate to cytoplasmic domain serine residues.
CC	-1- SIMILARITY: Contains 1 link domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M33827; AAA36967.1; .
DR	PIR; A35616; A35616.
DR	HSSP; P98066; ITSG.
DR	InterPro; IPRO01231; CD44_antigen.
DR	InterPro; IPRO00538; Link.
DR	Pfam; PF00193; XLink; 1.
DR	PRINTS; PR0658; CD44.
DR	PRINTS; PR01265; LINKMODULE.
DR	ProDom; PD000918; Link; 1.
DR	SMART; SMO0445; Link; 1.
KW	PROSITE; PS01241; Link; 1.
KW	Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
KW	Proteoglycan; Signal; Alternative splicing;
KW	Pyroglutamate carboxylic acid.
KW	SIGNAL
FT	1 22
FT	CHAIN
FT	DOMAIN 23 362
FT	TRANSMEM 23 269
FT	DOMAIN 291 290
FT	DOMAIN 291 362
FT	DOMAIN 50 121
FT	DOMAIN 152 160
FT	DOMAIN 226 269
FT	MOD_RES 23 23
FT	DISULFID 55 120
FT	DISULFID 79 99
FT	MOD_RES 292 292
FT	MOD_RES 326 326
FT	CARBOHYD 27 27
FT	CARBOHYD 59 59
FT	CARBOHYD 102 102
FT	CARBOHYD 112 112
FT	CARBOHYD 122 122
FT	CARBOHYD 174 174
FT	CARBOHYD 256 256
SEQ	SEQUENCE 362 AA; 39775 MW; ESBE9B34EBEBC948 CRC64;

Query Match 53.6%; Score 1398.5; DB 1; Length 362;
 Best Local Similarity 58.4%; Pred. No. 3.5e-80;
 Matches 291; Conservative 26; Mismatches 40; Indels 141; Gaps 7;

QY 1 MOKFMWAMGLCYPLSLA--QIDILNTRCAGVFEHKKGRYSISRTPEADCKAPNS 58
 1 MOKFMWAMGLCYPLSLA--QIDILNTRCAGVFEHKKGRYSISRTPEADCKAPNS 58
 DB 1 MOKFMWAMGLCYPLSLA--QIDILNTRCAGVFEHKKGRYSISRTPEADCKAPNS 60
 QY 59 TLPTAOMKALISGFTCRGYPFEGHVVIPRIHNSICANNVTGYLLTNTSGYDTYC 118
 DB 61 TLPTAOMKALISGFTCRGYPFEGHVVIPRIHNSICANNVTGYLLTNTSGYDTYC 120
 QY 119 FNASAPREDCTSVNDLPNAPDGLTITTVNDGTRVYKGRFRTNPEDIPSNPTDDV 178
 DB 121 FNASAPREDCTSVNDLPNAPDGLTITTVNDGTRVYKGRFRTNPEDIPSNPTDDV 180
 QY 179 SSGSSRSSTSGYTFYTF-STVHPIDPDSFWITDSTRIPATNDSHSTTLQPTAN 237
 DB 181 SSGSSRSSTSGYTFYTF-STVHPIDPDSFWITDSTRIPATNDSHSTTLQPTAN 218
 QY 238 PNTGLVEDLDRTGPIASMTQOQNSQSFSTSHGLEDHDPPTSTLTSNRDVTGERD 297
 DB 219 -----MATR----- 222
 QY 298 PNHSEGSTLLEGYTSHYPTKESRTFIPVTSKATGSFGVATVVGDSNVNRSLSGDQ 357
 DB 223 -----DQ 224
 QY 358 D-TFHPGSGSH-TTHGSEBDGSHSQSGCANNTGSPITTPQIPFWLLILSLALAIL 415
 DB 225 DSDMDPRGNSLFTVTDSSKLTSHSSGQDGLNSTSPGPKPRVPMWLLVLSLALAIL 284
 QY 416 AVGIAVNSRRCCQKKLVINSNGAVNDEPKSPGNGEASKEQEWLVANKESSPTPOF 475
 DB 285 AVGIAVNSRRCCQKKLVINSNGAVNDEPKSPGNGEASKEQEWLVANKESSPTPOF 344
 QY 476 MTADEFRLQNVDMKIGV 493
 DB 345 MTADEFRLQNVDMKIGV 362

RESULT 10
 SERI_BOMMO STANDARD; PRT; 389 AA.
 AC P07856;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sericin precursor (Silk gum protein).
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Middle silk gland;
 RX MEDLINE=83082839; PubMed=6294094;
 RA Okamoto H., Ishikawa S., Suzuki Y.,
 RA "Structural analysis of sericin genes. Homologies with fibroin gene
 RA in the 5' flanking nucleotide sequences."
 RT J. Biol. Chem. 257.15192-15199 (1982).
 RN [2]
 RP SEQUENCE OF 317-354 FROM N.A.
 RX MEDLINE=87076763; PubMed=3024742;
 RA Michaille J.-C., Goude P., Prudhomme J.-C., Garel A.;
 RA "A single gene produces multiple sericin messenger RNAs in the silk
 RA gland of Bombyx mori."
 RT Biochimie 68:1165-1173 (1986).
 CC -1- FUNCTION: Provides the silk fibroin thread with a sticky coating.
 CC Acts as a cement by sticking silk threads together.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Produced exclusively in the middle (MSG)

CC section of silk glands.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

CC -----
 CC EMBL; J01029; NOT ANNOTATED CDS.
 CC EMBL; J01030; NOT ANNOTATED CDS.
 CC EMBL; J01031; NOT ANNOTATED CDS.
 CC EMBL; J01032; NOT ANNOTATED CDS.
 CC EMBL; J01033; NOT ANNOTATED CDS.
 CC EMBL; J01034; NOT ANNOTATED CDS.
 CC EMBL; M26101; AAA27843.1; --
 CC EMBL; M26102; AAA27844.1; --
 CC SIK; Signal.
 CC SIGNAL 1 15 POTENTIAL.
 CC CHAIN 16 389 SERICIN.
 CC SEQUENCE 389 AA; 38820 MW; BA605BC0305EAF19 CRC64;

Query Match 6.8%; Score 177; DB 1; Length 389;
 Best Local Similarity 23.8%; Pred. No. 0.00036;
 Matches 101; Conservative 55; Mismatches 131; Indels 138; Gaps 18;

QY 38 KNGRYSISRTPEADLCFAFNSTLPTAOMKAL-----SIGFET 76
 DB 32 KNRKTAASSESSYLND-NDISAGAHRAKSYGQODSKYTSRGEVYSGRQNYKD 90
 QY 77 CRYPFIEGHVPIRIHNSICANNVTGYLLTNTGYD-TYFNASAPREDCTSVTL 135
 DB 91 SKQALLISGCT---KSSNSVQSDSKSAQSSSSRSQESAYSSSSSSSTEBSSSSRA 147
 QY 136 PNAFDGPITLTVNRDSTRVYKGRFRTNPEDIPSNPTDDVVS--GSSSER----- 186
 DB 148 ASSTDAS-SNTDSNSAGSSTGGRTYV---YSSNSRGSVSSSTSSNTDSNSNAG 203
 QY 187 SSTSGGYLYFTFTVHPIDPDSFWITDSTRIPATNDSHSTTLQPTANPTGLVEDL 246
 DB 204 SSTSGGSSTGYVS-----SNRDSVST----- 226
 QY 247 DRGPIASMTQOQNS-----QSPSTSHGLEDHDPPTSTLTSNR-----NDVTG 293
 DB 227 --TSSSNTDSNSNSVGSRRSGSSSHEDSKSDENVSTTSSSNTDSNSVGSSTG 284
 QY 294 GRR-----DPN-HSEGSTLLEGYTSHYPTKESRTFIPVTS 330
 DB 285 GRRTYGVSSNRDQSVSTGSSSNTDSNSVGSST--SGGSSTYGVSSNSR--DGSVS 339
 QY 331 KTSFGYATVAVGDSNNVNRSLSGDDPTHPGGSSTTGSSSDGSHSQSGCANNTS 390
 DB 340 STGSSSNT-----DSNSN-----SAGSSTSGGSSSTGYSSNSHDSVSTG 380
 QY 391 GPIRT 395
 DB 381 SSSNT 385

RESULT 11
 TSG6_RABIT STANDARD; PRT; 276 AA.
 AC P98065;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tumor necrosis factor-inducible protein TSG-6 precursor (TNF-
 DE stimulated gene 6 protein) (hyaluronate-binding protein P64).
 GN TNFIP6 OR TSG6 OR P64.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

CC NCBI_TaxID=9986;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=New Zealand white;
 CC MEDLINE=93252803; PubMed=8098034;
 CC Feng P., Liu G.;
 CC "Identification of a novel serum and growth factor-inducible gene in
 CC vascular smooth muscle cells."
 CC J. Biol. Chem. 268:9387-9392(1993).
 CC [2]
 CC ERRATUM.
 CC MEDLINE=94012707; PubMed=8407990;
 CC Feng P., Liu G.;
 CC J. Biol. Chem. 268:21453-21453(1993).
 CC -1- FUNCTION: Possibly involved in cell-cell and cell-matrix
 CC interactions during inflammation and tumorigenesis.
 CC -1- TISSUE SPECIFICITY: Vascular smooth muscle cells.
 CC -1- DEVELOPMENTAL STAGE: Fetal.
 CC -1- INDUCTION: By serum and growth factor.
 CC -1- SIMILARITY: Contains 1 link domain.
 CC -1- SIMILARITY: Contains 1 CUB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M86381; AAA03342.1; -
 CC PIR: A48055; A47290.
 CC HSP: P98066; ITSG.
 CC InterPro: IPR000859; CUB.
 CC InterPro: IPR00538; Link.
 CC Pfam: PF00431; CUB; 1.
 CC Pfam: PF00193; Xlink; 1.
 CC PRINTS: PRO1265; LINKMODULE.
 CC ProDom: PD000918; Link; 1.
 CC SMART: SMO0042; CUB; 1.
 CC SMART: SMO0445; LINK; 1.
 CC PROSITE: PS01180; CUB; 1.
 CC PROSITE: PS01241; LINK; 1.
 CC Cell adhesion; Signal; Glycoprotein.
 CC SIGNAL 1
 CC CHAIN 20 276 POTENTIAL.
 CC TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN
 CC TSG-6.
 CC LINK.
 CC CUB.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 276 AA; 31081 MW; 38C5D9A24B2F75A CRC64;
 CC SQ

Query Match 6.6%; Score 171; DB 1; Length 276;
 Best Local Similarity 27.2%; Pred. No. 0.00055;
 Matches 62; Conservative 34; Mismatches 64; Indels 68; Gaps 13;

QY 31 AGGVHVE-KNGRYS-SRTEDADLCAPFSTPLTMAQWKEKALSGFETCRVFI-EGHVI 88
 DB 35 AGVHREARSGKIKLTYAKKAVCFEGRLATYQLAAKIKGHVCAAGMAKGRVGY 94
 QY 89 PRIHNSICAPANNVTGY--ILTSNTGYDYTCFNASAPPEEDCTV-TD-----L 135
 DB 95 PIYKPGSNGCFGKTIHIVGILNRSERWDAYCYN--PKAKCGGVTFDKRIKFSRGF 151
 QY 136 PNAFPG-----PITITVARD-----GTRYVO-----KGEYRT 163
 DB 152 FMYEDNQCICWHIRLKIGQRHLISFLINFDLEYDPCCLADYVEIYSDVDVHGVRICG 211

QY 164 N--PEDITYPSNPDDVSSGS-----SERSSTGGYFYFTSTVHP 203
 DB 212 DELPEDI-----ISTENVMTLKEPLSDASVYAGGFQI-KYIVTDP 249

RESULT 12
 PCQA RAT STANDARD; PRT; 2124 AA.
 ID PCQA RAT
 AC P07897;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
 DE protein) (CSPCP).
 GN AGC1 OR AGC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=86087070; PubMed=3693370;
 CC Doege K., Saaki M., Horigan E., Hassell J.R., Yamada Y.;
 CC "Complete primary structure of the rat cartilage proteoglycan core
 CC protein deduced from cDNA clones."
 CC J. Biol. Chem. 262:17757-17767(1987).
 CC [2]
 CC REVISION TO 698.
 CC Doege K., Saaki M., Horigan E., Hassell J.R., Yamada Y.;
 CC J. Biol. Chem. 263:10040-10040(1988).
 CC [3]
 CC SEQUENCE OF 1856-2124 FROM N.A.
 CC MEDLINE=86250698; PubMed=242493;
 CC Doege K., Fernandez P., Hassell J.R., Saaki M., Yamada Y.;
 CC "Partial cDNA sequence encoding a globular domain at the C terminus
 CC of the rat cartilage proteoglycan."
 CC J. Biol. Chem. 261:8108-8111(1986).
 CC -1- FUNCTION: This proteoglycan is a major component of extracellular
 CC matrix of cartilaginous tissues. A major function of this protein
 CC is to resist compression in cartilage. It binds avidly to
 CC hyaluronic acid via an amino-terminal globular region. May play a
 CC regulatory role in the matrix assembly of the cartilage.
 CC -1- SUBUNIT: Interacts with PHLN (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
 CC similarity).
 CC -1- DOMAIN: Two globular domains, G1 and G2, comprise the amino
 CC terminus of the proteoglycan, while another globular region, G3,
 CC makes up the COOH terminus. G1 contains link domains and thus
 CC consists of three disulfide-bonded loop structures designated as
 CC the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)
 CC and the chondroitin sulfate (CS) attachment domains lie between G2
 CC and G3.
 CC -1- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate
 CC chains. N-linked and O-linked oligosaccharides.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 4 link domains.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13518; AAA1836.1; -
 CC EMBL: J03485; AAA21000.1; ALT SEQ.
 CC PIR: A92623; A28452.
 CC HSP: P98066; ITSG.
 CC InterPro: IPR002353; Antifreeze1.

CC or send an email to license@ib-sib.ch.

CC EMBL; U83903; AAC3527.1; -

DR PIR; J06506; J06506.

DR HESP; P98066; TSG.

DR MGI; M1195266; Tnfaiip6.

DR InterPro; IPR000859; CUB.

DR InterPro; IPR000538; Link.

DR Pfam; PF00431; CUB; 1.

DR Pfam; PF00193; XLink; 1.

DR PRINTS; PR01265; LINKMODULE.

DR ProDom; PD000918; Link; 1.

DR SMART; SM00042; CUB; 1.

DR SMART; SM00445; Link; 1.

DR PROSITE; PS01180; CUB; 1.

DR PROSITE; PS01241; Link; 1.

DR Cell adhesion; Signal; Glycoprotein.

KW SIGNAL 1 17 POTENTIAL.

FT CHAIN 18 275 TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN

FT DOMAIN 35 128 TSG-6.

FT DOMAIN 135 247 CUB.

FT DISULFID 58 127 BY SIMILARITY.

FT DISULFID 82 103 BY SIMILARITY.

FT DISULFID 135 161 BY SIMILARITY.

FT DISULFID 188 210 BY SIMILARITY.

FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 275 AA; 30924 MW; 1CD24728208BF9 CRC64;

Query Match 6.2%; Score 163; DB 1; Length 275;

Best Local Similarity 26.1%; Pred. No. 0.0017;

Matches 63; Conservative 34; Mismatches 76; Indels 68; Gaps 13;

QY 31 AGVFHVE-KNGRYSISRTADCKAFNSTLPTMAQMEKALSIGFETCYGPT-BGHVYI 88

DB 35 AGVYHRAARAGRKLTVAEKAVCEPFGGLATYKGLDEARKIGFHVCAAGMAYGRVGY 94

QY 89 PRHENSICANNVTGV--ILTSNTSQDYTCFNASAPPEEDCTSY-TD-----L 135

DB 95 PIYKPGPCGFRKGTGIIIDVGRINRSERWDAYCYN--PRAKCGGVFTDPRIKPSGPF 151

QY 136 PNAFPG-----PITTVNRDGT-----YVQ-----KGEYRT 163

DB 152 PNEYDNGVCYWHIRLKYGRILHSFLDPLEHDPGLADYVAIYVSVDYHGFRVYCG 211

QY 164 N--PEDIYPSNPITDDVSSGS-----SSERSSTSGGYFFYFSTYHPIDEDSPWITDS 215

DB 212 DELPEDI-----ISTGNVMTLKFSLDASYTAGGFQI-KYVTVDPASKSQANNTST 261

QY 216 T 216

DB 262 T 262

RESULT 14

TSG6 HUMAN

ID TSG6 HUMAN STANDARD; PRT; 277 AA.

AC P98066; O8MW19;

DT 01-FEB-1986 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Tumor necrosis factor-inducible protein TSG-6 precursor (TNF-stimulated gene 6 protein) (Hyaluronate-binding protein).

GN TNFAIP6 OR TSG6.

OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fibroblast;

RX MEDLINE=92112993; PubMed=1730767;

RA Lee T.H., Wisniewski H.-G., Vilcek J.

RT "A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44."

RT J. Cell Biol. 116:545-557(1992).

RL [2]

RP SEQUENCE FROM N.A., AND VARIANT GLN-144.

RX MEDLINE=21975206; PubMed=11854277;

RA Nentwich H.A., Mustafa Z., Rugg M.S., Marsden B.D., Cordell M.R., Mahoney D.J., Jenkins S.C., Dowling B., Fries E., Milner C.M., Loughlin J., Day A.J.

RT "A novel allelic variant of the human TSG-6 gene encoding an amino acid difference in the CUB module. Chromosomal localization, frequency analysis, modeling, and expression."

RT J. Biol. Chem. 277:15354-15362(2002).

RL [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung, and Spleen;

RX MEDLINE=22388257; PubMed=12477932;

RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.J., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Hulik S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butcherfield Y.S.N., Krzywinski M.I., Skalka U., Smalins D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RT Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).

RL [4]

RP STRUCTURE BY NMR OF 36-133.

RX MEDLINE=96390850; PubMed=8797823;

RA Kohda D., Morton C.J., Parkar A.A., Hatanaka H., Inagaki F.M., Campbell I.D., Day A.J.

RT "Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration."

RT Cell 86:767-775(1996).

RL -1- FUNCTION: Possibly involved in cell-cell and cell-matrix interactions during inflammation and tumorigenesis.

CC -1- INDUCTION: By TNF.

CC -1- SIMILARITY: Contains 1 CUB domain.

CC -1- SIMILARITY: Contains 1 CUB domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ihb-sib.ch/announce/> or send an email to license@ib-sib.ch).

CC EMBL; M31165; AB00792.1; -

DR EMBL; AJ421518; CAD13434.1; -

DR EMBL; AJ419936; CAD12353.1; -

DR EMBL; BC030205; AAH30205.1; -

DR PIR; A41735; A41735.

DR PDB; 1O7B; 07-NOV-03.

DR Genew; HGNC:11898; TNFAIP6.

DR MIM; 600410;

DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.

DR GO; GO:0005540; F:hyaluronic acid binding; TAS.

DR GO; GO:0007267; P:cell-cell signaling; TAS.

DR GO; GO:0006954; P:inflammatory response; TAS.

GO; Co:0007165; P:signal transduction; TAS.

DR InterPro: IPR000859; CUB.

DR InterPro: IPR000538; Link.

DR Pfam: PF00431; CUB; 1.

DR Pfam: PF00193; Xlink; 1.

DR PRINTS; PR01265; LINKMODULE.

DR ProDom; PD000918; Link; 1.

DR SMART; SM00042; CUB; 1.

DR SMART; SM00445; LINK; 1.

DR PROSITE; PS01180; CUB; 1.

DR PROSITE; PS01241; LINK; 1.

KW Cell adhesion; Signal; Glycoprotein; Polymorphism; 3D-structure.

FT SIGNAL 1 19

FT CHAIN 20 277

FT

FT DOMAIN 53 128

FT DOMAIN 135 247

FT DISULFID 58 127

FT DISULFID 82 103

FT DISULFID 135 161

FT DISULFID 188 210

FT CARBOHYD 118 118

FT CARBOHYD 258 258

FT VARIANT 144 144

FT STRAND 38 39

FT HELIX 51 60

FT TURN 71 72

FT STRAND 85 85

FT STRAND 88 88

FT TURN 89 90

FT STRAND 91 91

FT TURN 103 106

FT TURN 116 117

FT STRAND 124 124

FT STRAND 126 127

SO SEQUENCE 277 AA; 31231 MW; 4DB3AEB4AC52B880 CRC64;

Query Match 6.2%; Score 162.5; DB 1; Length 277;

Best local Similarity 33.3%; Pred. No. 0.0019;

Matches 40; Conservative 21; Mismatches 50; Indels 9; Gaps 5;

29 RFAGVPHVE-KNGRYSISTEADLCRAFNSTLPTMNMOMKALSIGPGRGFEI-EGHY 86

33 RAAGVTHREARSKRYLTIAEKAVCEFGCHLTITYKLEBARLIGFHVCAAMAKGRV 92

87 VPIRIHPSNICAAANTGVY--ILTSNTSQYDYTCFNASAPPEEDCTSV-TDLPNADGP 142

93 GYPIVAPGPNCGCGKGTGLIDYIGIRINRSEKWDAYCYN--PFAKECGVFTDPKRIFKSP 149

RESULT 15

MUC1_MOUSE STANDARD; PRT; 630 AA.

AC Q02436;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Epistatin).

GN MUC1 OR MUC-1

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

11

RP SEQUENCE FROM N.A.

RX MEDLINE=91332029; PubMed=1714452; Splicier A.P., Parry G., Patton S.S.;

RT "Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism.";

RT J. Biol. Chem. 266:15099-15109(1991).

[illegible]

FT	CONFLICT	121	129	Lt -> S (IN REF. 2).
FT	CONFLICT	138	131	AT -> PA (IN REF. 2).
FT	CONFLICT	140	140	T -> TT (IN REF. 2).
FT	CONFLICT	423	423	F -> S (IN REF. 2).
FT	CONFLICT	506	506	S -> D (IN REF. 2).
FT	CONFLICT	602	602	Q -> S (IN REF. 2).
SQ	SEQUENCE	630 AA:	64622 MW;	FEF7C1B3137F683B CRC64

Query Match	6.18;	Score 158;	DB 1;	Length 630;
Best local similarity	36.34;	Prod No. 0 01;		

Best Local Similarity 26.2%; Pred. No. 0.01;
Matches 84; Conservative 42; Mismatches 136; Indels 58; Gaps 15;

```

Qy      8  IEGHVVIPRHPMSICAMNTGVILITNSTSGDYDPCFNAAGP----- 125
      18  LKGFLLALPS-EENSIVTSSQDTSSSLASTTTPVHSSNDPAIRPQDGTSSPVQSTSSPA 76
Db
Qy      126  ---EEDCTSYTDLPAADFPIITITLVNRDGRVYQKGRYRNPEDVIPSNTDDVSSGS 182
      77  TRAPEOSTS-TALSGTSSPATTAPN-SASSPAVHG-----TSSPATSPLDSN 125
Db
Qy      183  SSR-SRSTGCGYIFTFSTVAPHPEDSPWITSDTRIPATNM--DSHSTLTQTPANP 238
      126  SSPVYHSGTSSA-----ATTAPVGTSSPVHGGTSS-PATSPQDGTSSPDHSTSSP 178
Db
Qy      239  NTGLVEDLDTGTLSTMTQOANSQSF-STSHSGLEBEDDHTTSTLSSNRNDVT-CGR 236
      179  ATRAPESGTSTAVLSGTSSPATTAPVDSTSSVAAHDSTSPATSISSEDSASPVAHGQTS 238
Db
Qy      297  DPNHSEOSTILLEGVYSHYPHTMESRTFIPVTS-AKTSGFVAVTVGDSNWNANSLS 354
      239  SP-----ATSPFLNDSTSSPVHSSASLQIKTSTDLASTPDHNGTSTVTT--TSSALGSATS 291
Qy      355  GDQDTFRPSCGSHTTGSGS 374
Db      292  PDH-----SGTSTTNSSS 306

```

Search completed: March 8, 2004, 06:08:51
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 45 Seconds
(without alignments)
3456.677 Million cell updates/sec

Title: PCT-US01-51014-1

Sequence: 1 MDKFWHAAWGLCLVPLSLA.....QFMADETRLQNVDMKGV 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2608	100.0	493	4	Q86227
2	2490	95.5	493	4	Q96124
3	2466	94.6	470	4	Q9H5A6
4	2457.5	94.2	742	4	Q8U336
5	2348	90.0	676	4	Q9H5A7
6	2326.5	89.2	719	4	Q9H5A5
7	2189.5	84.0	514	11	Q62913
8	1837	70.4	361	4	Q8E772
9	1835	70.4	361	4	Q8N694
10	1748.5	67.0	580	11	Q80X37
11	1705	65.4	338	4	Q9H5A4
12	1637.5	62.8	780	11	Q08779
13	1536.5	58.9	364	6	Q97569
14	1494	57.3	294	4	Q92493
15	1367.5	52.4	364	11	Q70509
16	1358	52.1	271	4	Q9H5A3

17	871	33.4	177	4	Q99900	Q99900 homo sapien
18	836.5	32.1	398	13	Q90ZL8	Q90L18 anas platyr
19	804.5	30.6	396	13	Q9W6S4	Q9W6S4 gallus gall
20	662	25.4	127	4	Q8P222	Q9P222 homo sapien
21	646	24.8	183	6	Q9T0U9	Q9T0U9 bos taurus
22	638	24.5	183	6	Q9T0U8	Q9T0U8 bos indicus
23	634	24.3	183	6	Q9T0U5	Q9T0U5 bos taurus
24	502	19.2	265	13	Q98SR5	Q98SR5 anas platyr
25	450	17.3	168	13	Q90ZL6	Q90ZL6 anas platyr
26	432	16.6	106	5	Q17448	Q17448 strongyloce
27	422	16.2	437	11	Q35249	Q35249 rattus norv
28	417	16.0	80	4	Q86VZ1	Q86VZ1 homo sapien
29	417	16.0	139	4	Q9S370	Q9S370 homo sapien
30	200	7.7	318	11	Q8BHC0	Q8BHC0 mus musculu
31	200	7.7	318	11	Q99NE4	Q99NE4 mus musculu
32	199.5	7.6	322	4	Q8TC18	Q8TC18 homo sapien
33	199.5	7.6	322	4	Q9UNF4	Q9UNF4 homo sapien
34	192.5	7.4	332	4	Q9Y5Y7	Q9Y5Y7 homo sapien
35	185.5	7.1	162	11	Q9SRN2	Q9SRN2 rattus norv
36	183	7.0	222	5	Q8IFX6	Q8IFX6 caenorhabdi
37	178	6.8	833	16	Q98Q44	Q98Q44 mycoplasma
38	172	6.6	94	11	Q9ERN3	Q9ERN3 rattus norv
39	171	6.6	95	11	Q9EP99	Q9EP99 rattus norv
40	169	6.5	372	6	Q28285	Q28285 canis fami
41	167	6.4	2310	16	Q8CMU7	Q8CMU7 staphylococ
42	165	6.3	786	5	Q21027	Q21027 caenorhabdi
43	164.5	6.3	631	11	Q9YK60	Q9YK60 mus musculu
44	162.5	6.2	2772	5	Q9VAV4	Q9VAV4 drosophila
45	161	6.2	1275	5	Q76602	Q76602 caenorhabdi

ALIGNMENTS

RESULT 1

ID Q86227 PRELIMINARY; PRT; 493 AA.
AC Q86227;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD44 antigen.
GN CD44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP TISSUE=Colon adenocarcinoma;
RC Wiese G.J., Freund D., Corbell D.;
RA "Sequence analysis of the human CD44 antigen."
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY101192; AM50040.1; -;
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO: GO:0005540; F:hyaluronic acid binding; IEA.
DR GO: GO:0007155; F:cell adhesion; IEA.
DR InterPro: IPR001231; CD44 antigen.
DR InterPro: IPR000536; Link.
DR Pfam: PF00193; Xlink: 1.
DR PRINTS: PR00658; CD44.
DR PRINTS: PR01265; LINKMODULF.
DR PRODOM: PD000916; Link: 1.
DR SMART: SM00445; Link: 1.
DR PROSITE: PS01241; Link: 1.
SQ SEQUENCE 493 AA; 53398 MW; 9C59CCCC8C261BF6 CRC64;

Query Match 100.0%; Score 2608; DB 4; Length 493;
Best Local Similarity 100.0%; Pred. No. 3.9e-177; Indels 0; Gaps 0;
Matches 493; Conservative 0; Mismatches 0;

```

Db      1 MDKFWMAAMGLCLVPLSLAQIDPLNTCRFAGVFEKNGRYSISREADLCKAFNSTL 60
Qy      61 PTMAOMKALSIGETCRGYFIBGHVVIPIIHNSICANNNGVYILTSNTSOYDYTCFN 120
Db      61 PTMAOMKALSIGETCRGYFIBGHVVIPIIHNSICANNNGVYILTSNTSOYDYTCFN 120
Qy      121 ASAPPEEDCTSVTLPAFADGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180
Db      121 ASAPPEEDCTSVTLPAFADGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180
Qy      181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATMDSHSTTLOPTANPMT 240
Db      181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATMDSHSTTLOPTANPMT 240
Qy      241 GLVEDLDRTGPISLMTQOSNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300
Db      241 GLVEDLDRTGPISLMTQOSNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300
Qy      301 SEGSTLLEGYTSHPYPTKESRPIPVTSAKTSGFVTAATVGDNSNVNRSLSGDQTF 360
Db      301 SEGSTLLEGYTSHPYPTKESRPIPVTSAKTSGFVTAATVGDNSNVNRSLSGDQTF 360
Qy      361 HPSCGSHTTGSESDGSHSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420
Db      361 HPSCGSHTTGSESDGSHSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420
Qy      421 VNSRRRCQKKQVLVNSGNGAVEDRKPSGLNGEASKSQEMVHLVNKESSTPDQMTADE 480
Db      421 VNSRRRCQKKQVLVNSGNGAVEDRKPSGLNGEASKSQEMVHLVNKESSTPDQMTADE 480
Qy      481 TRATQNVDMKIGV 493
Db      481 TRATQNVDMKIGV 493

```

RESULT 2

```

ID      096J24      PRELIMINARY;      PRT;      699 AA.
AC      096J24      01-DEC-2001 (TREMblrel. 19, Created)
DT      01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT      01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE      Similar to CD44 antigen (Homing function and indian blood group
        system).
OS      Homo sapiens (Human).
CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RA      Strausberg R.;
RL      Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC004372; AA04372.1; -.
DR      PIR; I37369; I37369.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR      GO; GO:0005540; F:hyaluronic acid binding; IEA.
DR      GO; GO:0007155; P:cell adhesion; IEA.
DR      InterPro; IPR001231; CD44_antigen.
DR      InterPro; IPR000538; Link.
DR      Pfam; PF00193; Xlink; 1.
DR      PRINTS; PR00658; CD44.
DR      ProDom; PD00919; Link; 1.
DR      SMART; SM00445; Link; 1.
DR      ProSITE; PS01241; Link; 1.
SQ      SEQUENCE 699 AA; 76628 MW; 68AC366A92DBC7C CMC64;

```

Query Match 95.5%; Score 2490; DB 4; Length 699;
 Best Local Similarity 70.4%; Pred. No. 1.5e-168;
 Matches 492; Conservative 1; Mismatches 0; Indels 206; Gaps 1;

```

Qy      1 MDKFWMAAMGLCLVPLSLAQIDPLNTCRFAGVFEKNGRYSISREADLCKAFNSTL 60
Db      1 MDKFWMAAMGLCLVPLSLAQIDPLNTCRFAGVFEKNGRYSISREADLCKAFNSTL 60
Qy      61 PTMAOMKALSIGETCRGYFIBGHVVIPIIHNSICANNNGVYILTSNTSOYDYTCFN 120
Db      61 PTMAOMKALSIGETCRGYFIBGHVVIPIIHNSICANNNGVYILTSNTSOYDYTCFN 120
Qy      121 ASAPPEEDCTSVTLPAFADGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180
Db      121 ASAPPEEDCTSVTLPAFADGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180
Qy      181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATMDSHSTTLOPTANPMT 240
Db      181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATMDSHSTTLOPTANPMT 240
Qy      241 GLVEDLDRTGPISLMTQOSNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300
Db      241 GLVEDLDRTGPISLMTQOSNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300
Qy      301 SEGSTLLEGYTSHPYPTKESRPIPVTSAKTSGFVTAATVGDNSNVNRSLSGDQTF 360
Db      301 SEGSTLLEGYTSHPYPTKESRPIPVTSAKTSGFVTAATVGDNSNVNRSLSGDQTF 360
Qy      421 VNSRRRCQKKQVLVNSGNGAVEDRKPSGLNGEASKSQEMVHLVNKESSTPDQMTADE 480
Db      421 VNSRRRCQKKQVLVNSGNGAVEDRKPSGLNGEASKSQEMVHLVNKESSTPDQMTADE 480
Qy      481 TRATQNVDMKIGV 493
Db      481 TRATQNVDMKIGV 493

```

RESULT 3

```

ID      09H5A6      PRELIMINARY;      PRT;      470 AA.
AC      09H5A6      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT      01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE      Similar to CD44 antigen (Homing function and indian blood group
        system) (Fragment).
OS      Homo sapiens (Human).
CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      Cobley V.;
RA      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AL133330; CAC10346.1; -.
DR      HSSP; P98066; 1TSG.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR      GO; GO:0005540; F:hyaluronic acid binding; IEA.

```


DR GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRODOM; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 FT NON TER 1
 SQ SEQUENCE 470 AA; 50727 MW; E4B47CF12500D568 CRC64;

Query Match 94.6%; Score 2466; DB 4; Length 470;
 Best Local Similarity 99.8%; Pred. No. 4.5e-167;
 Matches 469; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DR 24 LITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTLPTMAQMEKALSIGFETCRYGFI 83
 Db 1 LITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTLPTMAQMEKALSIGFETCRYGFI 60

DR 84 GHVVIPIRHNSICANNNGVYILTSNTSYDTTCFNASAPPEEDCTSVTDLPNADGPI 143
 Db 61 GHVVIPIRHNSICANNNGVYILTSNTSYDTTCFNASAPPEEDCTSVTDLPNADGPI 120

DR 144 TITLVNRDGRVYQKGRYRNPEDIPSNPTDDVSGSSSESSSTSGYIFPTSTVAP 203
 Db 121 TITLVNRDGRVYQKGRYRNPEDIPSNPTDDVSGSSSESSSTSGYIFPTSTVAP 180

DR 204 IPEDSPWITDSTRIPATNMDSSHSTLQPTANPTGLVEDLDRTPLSMTQOSNSQS 263
 Db 181 IPEDSPWITDSTRIPATNMDSSHSTLQPTANPTGLVEDLDRTPLSMTQOSNSQS 240

DR 264 FSTSHBELDKDHPHTSTLTSSNRNDVYGGRRDPNHSSESTLLAGTSHYPTKESRT 323
 Db 241 FSTSHBELDKDHPHTSTLTSSNRNDVYGGRRDPNHSSESTLLAGTSHYPTKESRT 300

DR 324 FIPVTSKTSFGVTAVTVDSDNSNVNRSLSGDDTFHPSGSHTHGSESDGSHSQE 383
 Db 301 FIPVTSKTSFGVTAVTVDSDNSNVNRSLSGDDTFHPSGSHTHGSESDGSHSQE 360

DR 384 GGANTSGPRTTQIPFWLITLASLALAILAVCIANVSRRCGGKKLVINSNGAVE 443
 Db 361 GGANTSGPRTTQIPFWLITLASLALAILAVCIANVSRRCGGKKLVINSNGAVE 420

DR 444 DRKPSGLNGEASKSQEWHLVNKESETPDQFMTADETRNLQNVMDKIGV 493
 Db 421 DRKPSGLNGEASKSQEWHLVNKESETPDQFMTADETRNLQNVMDKIGV 470

RESULT 4
 Q9UJ36 PRELIMINARY; PRT; 742 AA.
 ID Q9UJ36
 AC Q9UJ36
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Transmembrane glycoprotein precursor.
 GN CD44.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gunther U.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94147793; Pubmed=7508842;
 RA Gunther U.;
 RT "CD44: a multitude of isoforms with diverse functions";
 RL Curr. Top. Microbiol. Immunol. 184:47-63 (1993).
 DR EMBL; AJ251595; CAB61878.1; -.

DR PIR; A47195; A47195.
 DR HSSP; P98066; ITSG.
 DR GO:0016020; C:membrane; IEA.
 DR GO:0004895; F:hyaluronic acid binding; IEA.
 DR GO:0005440; F:cell adhesion receptor activity; IEA.
 DR GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRODOM; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 FT SIGNAL 1
 FT CHAIN 24
 SQ SEQUENCE 742 AA; 81598 MW; C319E5CE50B51D3 CRC64;

Query Match 94.2%; Score 2457.5; DB 4; Length 742;
 Best Local Similarity 66.0%; Pred. No. 3.2e-166;
 Matches 490; Conservative 2; Mismatches 1; Indels 249; Gaps 1;

DR 1 MDKFWMAAGCLVPLSLAQIDLNTCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60
 Db 1 MDKFWMAAGCLVPLSLAQIDLNTCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60

DR 61 PTMAQMEKALSIGFETCRYGFI 120
 Db 61 PTMAQMEKALSIGFETCRYGFI 120

DR 121 ASAPPEEDCTSVTDLPNADGPI 180
 Db 121 ASAPPEEDCTSVTDLPNADGPI 180

DR 181 GSSSESSSTSGYIFPTSTVAP 222
 Db 181 GSSSESSSTSGYIFPTSTVAP 222

DR 241 AMDWFSMLPSESKNHLHTTQMAATSSNTISAGWEPNEBDEDRHLSPSSGGIDDD 300
 Db 241 AMDWFSMLPSESKNHLHTTQMAATSSNTISAGWEPNEBDEDRHLSPSSGGIDDD 300

DR 223 ----- 222
 Db 223 ----- 222

DR 301 EDLISITSTTPRAFDHTKQNDWTQNPBSHNPVLLQTTTWTVDVANGTAYEGNMN 360
 Db 301 EDLISITSTTPRAFDHTKQNDWTQNPBSHNPVLLQTTTWTVDVANGTAYEGNMN 360

DR 223 ----- 222
 Db 223 ----- 222

DR 361 PEAPPLIHHEHBEETPHSTSTIQATPSITEETATOKQWFGNWHGYNQTPREDS 420
 Db 223 ----- 222

DR 421 HSTTGTAASHTSHPMQGRTPSPEDDSMTDFNPISHPMGRGHQGRMDMDSSSTT 480
 Db 421 HSTTGTAASHTSHPMQGRTPSPEDDSMTDFNPISHPMGRGHQGRMDMDSSSTT 480

DR 232 LQPTANPTGLVEDLDRTPLSMTQOSNSQSGSTSHGLEDKHPHTSTLTSSNRNDV 291
 Db 232 LQPTANPTGLVEDLDRTPLSMTQOSNSQSGSTSHGLEDKHPHTSTLTSSNRNDV 291

DR 481 LQPTANPTGLVEDLDRTPLSMTQOSNSQSGSTSHGLEDKHPHTSTLTSSNRNDV 540
 Db 481 LQPTANPTGLVEDLDRTPLSMTQOSNSQSGSTSHGLEDKHPHTSTLTSSNRNDV 540

DR 292 TGGRRDPNHSSESTLLAGTSHYPTKESRTFPVTSKTSFGVTAVTVDSDNSNVN 351
 Db 541 TGGRRDPNHSSESTLLAGTSHYPTKESRTFPVTSKTSFGVTAVTVDSDNSNVN 600

DR 352 SLSDGDDTFHPSGSHTHGSESDGSHSQEGGANTTSGPIRTQIPFWLITLASLAL 411
 Db 601 SLSDGDDTFHPSGSHTHGSESDGSHSQEGGANTTSGPIRTQIPFWLITLASLAL 660

DR 412 ALTIANVCIAVNSRRCCGKKLVINSNGAVEDEKPSGLNGEASKSQEWHLVNKESEST 471
 Db 661 ALTIANVCIAVNSRRCCGKKLVINSNGAVEDEKPSGLNGEASKSQEWHLVNKESEST 720

DR 472 PDQFMTADETRNLQNVMDKIGV 493

Tue Mar 9 08:08:43 2004

pct-us01-51014-1.rpt

Page 4

Db 721 PDQFMTADETRNLQNVDMKIGV 742

RESULT 5

Q9H5A7 PRELIMINARY; PRT; 676 AA.
AC Q9H5A7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE D68D18.2.4 (CD4 antigen (Homing function and Indian blood group system)) (Fragment).
GN CD44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA COBLEY V.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133330; CAC10345.1; -.
DR HSSP; P98066; ITSG.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
DR GO; GO:0007155; F:cell adhesion; IEA.
DR InterPro; IPR001231; CD44_antigen.
DR Pfam; PF00193; Xlink; 1.
DR PRINTS; PR00658; CD44.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00445; Link; 1.
DR PROSITE; PS01241; Link; 1.
FT NON TER 1
SQ SEQUENCE 676 AA; 73929 MW; 122C7250B7DA0F47 CRC64;

Query Match 90.0%; Score 2348; DB 4; Length 676;
Best Local Similarity 69.2%; Pred. No. 1.7e-158; Indels 206; Gaps 1;
Matches 468; Conservative 1; Mismatches 1;

QY 24 LNTICRFAGVHVEKNGRYSISRTBADLCKAFNSTLPTVAQMEKALISIGETCRYGPIE 83
DB 1 LNTICRFAGVHVEKNGRYSISRTBADLCKAFNSTLPTVAQMEKALISIGETCRYGPIE 60
QY 84 GHVVIPIRIHNSICANNNGVYILTNTSOYDTYCFNASAPREEDCTSVTDLPNAPFGPI 143
DB 61 GHVVIPIRIHNSICANNNGVYILTNTSOYDTYCFNASAPREEDCTSVTDLPNAPFGPI 120
QY 144 TITVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFFTFSTVHP 203
DB 121 TITVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFFTFSTVHP 180
QY 204 IPDEDSPWITDSTRIPAT----- 222
DB 181 IPDEDSPWITDSTRIPATSTSSNTISAGWEPNEENEDERHLSFGSGIDDEDFISS 240
QY 223 ----- 222
DB 241 TISTTPRAFHTKQNDQWQNSPNSPEVLLQTTMTVDVDRNGTTAYEGNMPPEAHP 300
QY 223 ----- 222
DB 301 LIHHEHHEEETPHSTIGATPSSTEEATQKEQWGNRMHGYRQTPKEDSHSTTGT 360
QY 223 ----- 222
DB 361 AAASAHSHPWQRTTSPEDSWTFDFNPISHMGHQAQGRMDSDSHSLITQPTAN 420
QY 238 PNTGLVEDLRTGLSMWTOOSNQSFSTSHGEGLEDKDPFTTSTLTSNRANDVTGGRD 237
DB 421 PNTGLVEDLRTGLSMWTOOSNQSFSTSHGEGLEDKDPFTTSTLTSNRANDVTGGRD 480

QY 298 PNHSGSTLLLEGYTSHPHTKESRTFIPYTSANTGSGFYTAATVVGSDNSNVRSLSGDQ 357
DB 481 PNHSGSTLLLEGYTSHPHTKESRTFIPYTSANTGSGFYTAATVVGSDNSNVRSLSGDQ 540
QY 358 DTFHPGSGSHTHSGESDGHSHSQEGGANTTSGPIRTPOIPEMLITLASLALALILAY 417
DB 541 DTFHPGSGSHTHSGESDGHSHSQEGGANTTSGPIRTPOIPEMLITLASLALALILAY 600
QY 418 CIANRRRCGQKKLVYNSGNGAVEDRKPSGLNGEASKSQEYVHLVKNKESSETPDQFMT 477
DB 601 CIANRRRCGQKKLVYNSGNGAVEDRKPSGLNGEASKSQEYVHLVKNKESSETPDQFMT 660
QY 478 ADETRNLQNVDMKIGV 493
DB 661 ADETRNLQNVDMKIGV 676

RESULT 6

Q9H5A5 PRELIMINARY; PRT; 719 AA.
AC Q9H5A5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE D68D18.2.3 (CD4 antigen (Homing function and Indian blood group system)) (Fragment).
GN CD44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA COBLEY V.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133330; CAC10347.1; -.
DR HSSP; P98066; ITSG.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
DR GO; GO:0007155; F:cell adhesion; IEA.
DR InterPro; IPR001231; CD44_antigen.
DR Pfam; PF00193; Xlink; 1.
DR PRINTS; PR00658; CD44.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00445; Link; 1.
DR PROSITE; PS01241; Link; 1.
FT NON TER 1
SQ SEQUENCE 719 AA; 78854 MW; 36B747A896AD7380 CRC64;

Query Match 89.2%; Score 2326.5; DB 4; Length 719;
Best Local Similarity 65.1%; Pred. No. 6.3e-157; Indels 249; Gaps 1;
Matches 468; Conservative 1; Mismatches 1;

QY 24 LNTICRFAGVHVEKNGRYSISRTBADLCKAFNSTLPTVAQMEKALISIGETCRYGPIE 83
DB 1 LNTICRFAGVHVEKNGRYSISRTBADLCKAFNSTLPTVAQMEKALISIGETCRYGPIE 60
QY 84 GHVVIPIRIHNSICANNNGVYILTNTSOYDTYCFNASAPREEDCTSVTDLPNAPFGPI 143
DB 61 GHVVIPIRIHNSICANNNGVYILTNTSOYDTYCFNASAPREEDCTSVTDLPNAPFGPI 120
QY 144 TITVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFFTFSTVHP 203
DB 121 TITVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFFTFSTVHP 180
QY 204 IPDEDSPWITDSTRIPAT----- 222
DB 181 IPDEDSPWITDSTRIPATSTSSNTISATRTATKQETWDFSWLFLPSESKNLHTTTQ 240

QY 223 ----- 222
 Db 241 MACTSNTISAGMEPEBENEDERDLSTFGSGLDDEDFISSTISTTTPRADHTKONOD 300
 QY 223 ----- 222
 Db 301 MTQWNPESHNPVLLQTTTMTDNDVANGTTAYEGMNPAPPLIHEHEHEETPHSTS 360
 QY 223 ----- 222
 Db 361 TIOATSSSTTEETATQKQWFGNRMHEGRTQPKEDSHSTTGTAAASANTSHPMGRTP 420
 QY 223 ----- 254
 Db 421 SPEDSSWTFPFNISHPMGRGHOAGRMDSHS-TTQPTANPTGLVEDDRTGPLSM 480
 QY 255 TTQSNQSQSFSTSHGLEEDKHPTSTLTSSNRNDVYGRDPHSGSTLLGYSYSH 314
 Db 481 TTQSNQSQSFSTSHGLEEDKHPTSTLTSSNRNDVYGRDPHSGSTLLGYSYSH 540
 QY 315 YPTKESRTPIPTSAKTSFGVATVYGDNSNVNRSLSGQDTFHPGSGSHTHGSES 374
 Db 541 YPTKESRTPIPTSAKTSFGVATVYGDNSNVNRSLSGQDTFHPGSGSHTHGSES 600
 QY 375 DGHSHSGEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIANVRRRCGCKLIV 434
 Db 601 DGHSHSGEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIANVRRRCGCKLIV 660
 QY 435 INSGNCAVEDRKPGLNGEASKSQEWHLVYNKESSETPDQFTADETNLQNVDMKICV 493
 Db 661 INSGNCAVEDRKPGLNGEASKSQEWHLVYNKESSETPDQFTADETNLQNVDMKICV 719

RESULT 7

Q62913 PRELIMINARY; PRT; 514 AA.
 ID Q62913;
 AC Q62913;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE CD441 (Fragment).
 OS Rattus norvegicus (Rat).
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Chondrosarcoma;
 RA Stevens J.W., Midura R.J.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U46958; AAA92921.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRODOM; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 FT NON TER 1
 SQ SEQUENCE 514 AA; 55647 MW; 37A4984F0E3DDA92 CRC64;

Query Match 84.0%; Score 2189.5; DB 11; Length 514;
 Best Local Similarity 82.9%; Pred. No. 2, 2e-147;
 Matches 426; Conservative 1; Mismatches 0; Indels 87; Gaps 1;

QY 127 EDCTSYTDLPNAPDGPITITVNRDGRVYQKGEYRPNEDIYPSNPTDDVSSGSSSR 186
 Db 61 EDCTSYTDLPNAPDGPITITVNRDGRVYQKGEYRPNEDIYPSNPTDDVSSGSSSR 120
 QY 187 SSTSGGYIYYTFTVHPIPDSDPWITDSTDRIPAT----- 222
 Db 121 SSTSGGYIYYTFTVHPIPDSDPWITDSTDRIPATIQATPSSTTEETATQKQWFGNRM 180
 QY 223 ----- 222
 Db 181 HEGRTQPKEDSHSTTGTAAASANTSHPMGRTPPEBDSWTFPFNISHPMGRGHOAG 240
 QY 223 ----- 279
 Db 241 RMDMDSSHTTLOPTANPTGLVEDDRTGPLSMTTQSNQSQSFSTSHGLEEDKHPT 300
 QY 280 TSTLTSSNRNDVYGRDPHSGSTLLGYSYSHPTKESRTPIPTSAKTSFGVTA 339
 Db 301 TSTLTSSNRNDVYGRDPHSGSTLLGYSYSHPTKESRTPIPTSAKTSFGVTA 360
 QY 340 VTVGDNSNVNRSLSGQDTFHPGSGSHTHGSESDGSHSGOEGGANTTSGPIRTPOIP 399
 Db 361 VTVGDNSNVNRSLSGQDTFHPGSGSHTHGSESDGSHSGOEGGANTTSGPIRTPOIP 420
 QY 400 EWLIIILASLALAILAVCIANVRRRCGCKLIVNSGNCAVEDRKPGLNGEASKSQ 459
 Db 421 EWLIIILASLALAILAVCIANVRRRCGCKLIVNSGNCAVEDRKPGLNGEASKSQ 480
 QY 460 MVHLVYNKESSETPDQFTADETNLQNVDMKICV 493
 Db 481 MVHLVYNKESSETPDQFTADETNLQNVDMKICV 514

RESULT 8

Q86T72 PRELIMINARY; PRT; 361 AA.
 ID Q86T72;
 AC Q86T72;
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein DKFZP451K1918.
 GN DKFZP451K1918.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OX Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ansoerge W., Krieger S., Mewes H.W., Weill B., Amid C., Fobo G., Han M.,
 RA Oeanger A., Wiemann S.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL332642; CAD89965.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRODOM; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 361 AA; 39387 MW; F51A746B44387C54 CRC64;

Query Match 70.4%; Score 1837; DB 4; Length 361;
 Best Local Similarity 72.8%; Pred. No. 1, 6e-122;
 Matches 359; Conservative 0; Mismatches 2; Indels 132; Gaps 1;

Db 1 MDKFWHAAAGLCLVPLSLAQIDILNITCRAGVFHFKGRYSISRTAADLCKAFNSTL 60
QY PTAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Db 61 PTAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
QY 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRTRYOKGERTNPEDIPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRTRYOKGERTNPEDIPSNPTDDVSS 180
QY 181 GSSSERSTSGGYIFYFTFSTVHPPIPDSDSPMTDSTDRIPATMDSHSTTLQPTANPNT 240
Db 181 GSSSERSTSGGYIFYFTFSTVHPPIPDSDSPMTDSTDRIPATMDSHSTTLQPTANPNT 240
QY 241 GLVEDLDRTGRLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Db 241 GLVEDLDRTGRLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
QY 301 SEGSTTLLEGYTSHPHTKSRPIPVTSKATGSGVTAATVGDNSNVNRSLSGDDTF 360
Db 301 SEGSTTLLEGYTSHPHTKSRPIPVTSKATGSGVTAATVGDNSNVNRSLSGDDTF 360
QY 361 HPSGSHHTHSGESDGHSHGSEBAGANTSGPIRTPOIPFWLILASLALALILAVCIA 420
Db 229 HPSGSHHTHSGESDGHSHGSEBAGANTSGPIRTPOIPFWLILASLALALILAVCIA 420
QY 421 VNSRRRCGQKKKLYNSGNAVEDRKPSGLNGEASQSEVHLVKNKSESTPDQFMTADE 480
Db 289 VNSRRRCGQKKKLYNSGNAVEDRKPSGLNGEASQSEVHLVKNKSESTPDQFMTADE 480
QY 481 TRNLQNVDMKIGV 493
Db 349 TRNLQNVDMKIGV 361

RESULT 9

Q8N694 PRELIMINARY; PRT; 361 AA.

Q8N694; 01-OCT-2002 (Tremblrel. 22, Created)

Q8N694; 01-OCT-2002 (Tremblrel. 22, Last sequence update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8N694; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

QY 1 MDKFWHAAAGLCLVPLSLAQIDILNITCRAGVFHFKGRYSISRTAADLCKAFNSTL 60
Db 1 MDKFWHAAAGLCLVPLSLAQIDILNITCRAGVFHFKGRYSISRTAADLCKAFNSTL 60
QY 61 PTAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Db 61 PTAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
QY 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRTRYOKGERTNPEDIPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRTRYOKGERTNPEDIPSNPTDDVSS 180
QY 181 GSSSERSTSGGYIFYFTFSTVHPPIPDSDSPMTDSTDRIPATMDSHSTTLQPTANPNT 240
Db 181 GSSSERSTSGGYIFYFTFSTVHPPIPDSDSPMTDSTDRIPATMDSHSTTLQPTANPNT 240
QY 241 GLVEDLDRTGRLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Db 241 GLVEDLDRTGRLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
QY 301 SEGSTTLLEGYTSHPHTKSRPIPVTSKATGSGVTAATVGDNSNVNRSLSGDDTF 360
Db 301 SEGSTTLLEGYTSHPHTKSRPIPVTSKATGSGVTAATVGDNSNVNRSLSGDDTF 360
QY 361 HPSGSHHTHSGESDGHSHGSEBAGANTSGPIRTPOIPFWLILASLALALILAVCIA 420
Db 229 HPSGSHHTHSGESDGHSHGSEBAGANTSGPIRTPOIPFWLILASLALALILAVCIA 420
QY 421 VNSRRRCGQKKKLYNSGNAVEDRKPSGLNGEASQSEVHLVKNKSESTPDQFMTADE 480
Db 289 VNSRRRCGQKKKLYNSGNAVEDRKPSGLNGEASQSEVHLVKNKSESTPDQFMTADE 480
QY 481 TRNLQNVDMKIGV 493
Db 349 TRNLQNVDMKIGV 361

RESULT 10

Q8OX37 PRELIMINARY; PRT; 580 AA.

Q8OX37; 01-JUN-2003 (Tremblrel. 24, Created)

Q8OX37; 01-JUN-2003 (Tremblrel. 24, Last sequence update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Q8OX37; 01-OCT-2003 (Tremblrel. 25, Last annotation update)

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=salivary gland;
 RA Strauberg R.;
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC051388; AHS1388.1;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD4 antigen.
 DR InterPro; IPR00538; Link.
 DR Pfam; PF00193; Link.
 DR PRINTS; PR00658; CD4.
 DR PRODOM; PR01265; LINKMODULE.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 KR Hypothetical protein
 SQ SEQUENCE 580 AA; 63264 MW; F38A59C3AA8BA0F CRC64;
 Query Match 67.0%; Score 1748.5; DB 11; Length 580;
 Best Local Similarity 63.2%; Pred. No. 5.6e-116;
 Matches 368; Conservative 38; Mismatches 85; Indels 91; Gaps 10;
 QY 1 MDKRWMAAGLCIVPLSLA-----QIDNITCRFAGVHEKNGRYSISTEADLCXAF 56
 DB 1 MDKRWMAAGLCIVPLSLA-----QIDNITCRFAGVHEKNGRYSISTEADLCXAF 60
 QY 57 NSTLPTAOMEKALSIGETCRYGFIHGAVIPRIHNSICANNNTGYIL-TSNTSOYD 115
 DB 61 NSTLPTDOKMLAKSGFETCRYGFIHGAVIPRIHNSICANNNTGYILVTSNTHD 120
 QY 116 TYCENASAPREDCTSVTDLPNADFPTITTYNRDCTRYVQGEKRTNPEDIPENPFD 175
 DB 121 TYCENASAPREDCTSVTDLPNADFPTITTYNRDCTRYVQGEKRTNPEDIPENPFD 180
 QY 176 DVSSGSSERSSTSGGIFFYTF-----STVH----- 202
 DB 181 DVSSGSSERSSTSGGIFFYTF-----STVH----- 202
 QY 203 -----PIPEDS-----PWTSTORI----- 219
 DB 240 WFGQWQGNKPPTSESHVTEGTASAHNNHPEQRITTSQOEDVSV-TDFEPIHPMG 298
 QY 220 -----PATNDSSTLTLPANNTGLVEDLRTGFLSMTQOOSNGSFTSHEGLEED 274
 DB 299 QGHOTESKOTDSSTLTLPANNTGLVEDLRTGFLSMTQOOSNGSFTSHEGLEED 358
 QY 275 KDHFSTLTLSNNDVYTGRRDPNHSBSTLLGTYSHYPTKESRTFIPVTSATGS 334
 DB 359 ENHPTSTILPSSTSGKACARCGSLPTDTTVEGYTFQYDPMENGLTFPVTPATEV 418
 QY 335 FGVAIVYWG-DSNNVNRSLSGDDTFHSGGSH--TTHGSESDGSHSGOEGANTSG 391
 DB 419 FGEIVTLATDSNVVDSLPGRDSSGDSGSRITYTHGSELGHSANODSGVTTTSG 478
 QY 392 PIRTPQIPFWLIIASLALALILAVCIANVSRRCQKKLVINSNGAVEDEKPSGLN 451
 DB 479 PMRRQIPFWLIIASLALALILAVCIANVSRRCQKKLVINSNGAVEDEKPSGLN 538
 QY 452 GEASKEGMEVLYNKESETPDQFMATDETRNLQNVMKIGV 493
 DB 539 GEASKEGMEVLYNKESETPDQFMATDETRNLQNVMKIGV 580

DE DUG8D18.2.2 (CD44 antigen (Homing function and indian blood group
 DE system)) (Fragment).
 GN CD44.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cobley V.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL133330; CAC10348.1;
 DR HSSP; P98066; ITSG.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44 antigen.
 DR InterPro; IPR00538; Link.
 DR Pfam; PF00193; Link; 1.
 DR PRINTS; PR00658; CD4.
 DR PRODOM; PR01265; LINKMODULE.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 FT NON TER 1 1
 SQ SEQUENCE 338 AA; 36732 MW; 9835C309B62E4DB CRC64;
 Query Match 65.4%; Score 1705; DB 4; Length 338;
 Best Local Similarity 71.7%; Pred. No. 3.5e-113;
 Matches 337; Conservative 0; Mismatches 1; Indels 132; Gaps 1;
 QY 24 INITCRFAGVHEKNGRYSISTEADLCXAFNSTLPTAOMEKALSIGETCRYGFI 83
 DB 1 INITCRFAGVHEKNGRYSISTEADLCXAFNSTLPTAOMEKALSIGETCRYGFI 60
 QY 84 GHVVIPIHNSICANNNTGYILTSNTSOYDTCENASAPREDCTSVTDLPNADFPI 143
 DB 61 GHVVIPIHNSICANNNTGYILTSNTSOYDTCENASAPREDCTSVTDLPNADFPI 120
 QY 144 TITTYNRDCTRYVQGEKRTNPEDIPENPDDVSSGSSERSSTSGGIFFYFSTVHP 203
 DB 121 TITTYNRDCTRYVQGEKRTNPEDIPENPDDVSSGSSERSSTSGGIFFYFSTVHP 180
 QY 204 IPEDSPWITDSTDRIPATNDSSTLTLPANNTGLVEDLRTGFLSMTQOOSNGS 263
 DB 181 IPEDSPWITDSTDRIPATNDSSTLTLPANNTGLVEDLRTGFLSMTQOOSNGS 200
 QY 264 FETSHGLEEDMDHPTTSLTSNNDVYTGRRDPNHSBSTLLGTYSHYPTKESRT 323
 DB 201 ----- 200
 QY 324 FIPVTSATGSFGVAVTVGDSNVNRSLSGDDTFHSGGSHTHGSESDGSHSGOE 383
 DB 201 -----DQTFHPSGSGSHTHGSESDGSHSGOE 228
 QY 384 GGANTTSGPIRTFQIPFWLIIASLALALILAVCIANVSRRCQKKLVINSNGAVE 443
 DB 229 GGANTTSGPIRTFQIPFWLIIASLALALILAVCIANVSRRCQKKLVINSNGAVE 288
 QY 444 DRKPSGLNGEASKQEMVLYNKESETPDQFMATDETRNLQNVMKIGV 493
 DB 289 DRKPSGLNGEASKQEMVLYNKESETPDQFMATDETRNLQNVMKIGV 338

RESULT 11
 O9H5A4 PRELIMINARY; PRT; 338 AA.
 AC O9H5A4;
 DT 01-VAR-2001 (TRENBLrel. 16, Created)
 DT 01-VAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

RESULT 12
 O08779 PRELIMINARY; PRT; 780 AA.
 AC O08779;
 DT 01-JUL-1997 (TRENBLrel. 04, Created)
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE CD44 protein.

GN CD44.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BX; TISSUE=Pancreas;
 RA Holmann M.;
 RT "Rattus norvegicus CD44 protein sequence."
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U96138; AB54002.1; -.
 DR HSSP; P98066; ITRG.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44_antigen.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 DR SEQUENCE 780 AA; 85917 MW; CC4D35AB1EA7377C CRC64;

Query Match 62.8%; Score 1637.5; DB 11; Length 780;
 Best Local Similarity 47.2%; Pred. No. 6.3e-108; Indels 295; Gaps 12;
 Matches 370; Conservative 35; Mismatches 84;

QY 1 MDKFWMAWAG-LCLVPLSLA-QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 57
 DB 1 MDKFWMAWAGLCLVPLSLAQOQIDNITCRFAGVHVEKNGRYSISRTAADLCEAFN 60
 QY 58 STPLTMAOMEKALSIGFETCRGFIIEGHVVIPIRHPSICANNVTGYI-LTNTSOYDT 116
 DB 61 TLLPLTMAOMELAKRGFETCRGFIIEGHVVIPIRHPSICANNVTGYIILLASNTSHYDT 120
 QY 117 YCFNAPAPBEDCTSVTDLPNAPFDGPIITITIVNBDGTRYVOKGEYRNPEDIVPSNPTD 176
 DB 121 YCFNAPAPBEDCTSVTDLPNAPFDGPIITITIVNBDGTRYVOKGEYRNPEDIVPSNPTD 180
 QY 177 DVSSGSSSEBSSTSGGYIFYT-----STVHPIDED 208
 DB 181 DVSSGSSSEBSSTSGGYIFYT-----STVHPIDED 208
 QY 209 SPWV----- 212
 DB 240 NNMWISWFGNSGTTQTQDPTTATTAATMTPTPPKQEAQNWFSMFQPSKSHLH 299
 QY 213 -----TDSIDRIP----- 220
 DB 300 TTTKAPGTESNTNPTGMRKNEBNEBTDKYPNFGSGGIDDEDPISSITATTWVSAHTK 359
 QY 221 -----ATNMD----- 225
 DB 360 QNERTQWNPILHSNPEVLLQTTTRMTDIDRNSTAHGNTWQEPQPPFNHBYQDEETP 419
 QY 226 ----- 225
 DB 420 HATSTTWADPNSTTEBAATOKKWFENWOGKNPPTPSDSHTEGTTASAHNNHPSQRM 479
 QY 226 -----SSHTTLLQPTANPNTGLVEDLDRGTPL 252
 DB 480 TTQSGEDVSWTDFFDPIGHPMGQGHQTESKDTGSSHTTLLQPTANPNTGLVEDLDRGTPL 539
 QY 253 SNTTQGSNSQSFSTSHGLEDKDPHTTSTLTSNRKNDVYTGKRDPNHBSGSTLLAGYT 312
 DB 540 SVTTPQSHQNSFTLPGLEBEGEDHPHTTSVLPSSYK--SGRRRGSGLPDPTTSLGTYT 596
 QY 313 SHYPTKESRTPIPTSAKTSFGVTAATVG-DSNSNVRSLSGDQD-TFAPSGASHT-T 369

DB 597 PÖYEDIMENGTLPFVTPAKTEVFGEETGVTATDSNFVDSGLPDGDS SMDPGGFPTV 656
 QY 370 HGSSEDSHSGSGGEGGANTTSGPRTTQIPEMTILLASLALAILAVCIANVSRRCQ 429
 DB 657 HGSSEDSHSGSGGEGGANTTSGPRTTQIPEMTILLASLALAILAVCIANVSRRCQ 716
 QY 430 KKKLVINSNGAVDRKPSGCIANGASKSQEMVHLVNESSETPDQFMTADETNTQAVDM 489
 DB 717 KKKLVINSNGAVDRKPSGCIANGASKSQEMVHLVNESSETPDQFMTADETNTQAVDM 776
 QY 490 KIGV 493
 DB 777 KIGV 780

RESULT 13

097569 PRELIMINARY; PRT; 364 AA.
 ID 097569;
 AC 097569;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE CD44 (Fragment).
 OS Ceratotherium simum simum.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Ceratotherium.
 NCBI_TaxID=73337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Worley M.B.;
 RT "White rhinoceros CD44."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF045933; AAD02418.1; -.
 DR HSSP; P98066; ITRG.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR001231; CD44_antigen.
 DR InterPro; IPR000538; Link.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR00658; CD44.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 1.
 DR SMART; SM00445; Link; 1.
 DR PROSITE; PS01241; Link; 1.
 DR NON_TER 364
 FT SEQUENCE 364 AA; 39670 MW; C56B5E410AABICFP CRC64;

Query Match 58.9%; Score 1536.5; DB 6; Length 364;
 Best Local Similarity 64.1%; Pred. No. 3.6e-101;
 Matches 318; Conservative 17; Mismatches 26; Indels 135; Gaps 8;

QY 1 MDKFWMAWAGLCLVPLSLAQOQIDNITCRFAGVHVEKNGRYSISRTAADLCKAFN 60
 DB 1 MDKFWMAWAGLCLVPLSLAQOQIDNITCRFAGVHVEKNGRYSISRTAADLCEAFN 60
 QY 61 PTMAQWEKALSIGFETCRGFIIEGHVVIPIRHPSICANNVTGYIILTSNTSOYDTYCN 120
 DB 61 PTMAQWEKALSIGFETCRGFIIEGHVVIPIRHPSICANNVTGYIILTSNTSOYDTYCN 120
 QY 121 ASAPP-EEDCTSVTDLPNAPFDGPIITITIVNBDGTRYVOKGEYRNPEDIVPSNPTD 179
 DB 121 ASAPP-EEDCTSVTDLPNAPFDGPIITITIVNBDGTRYVOKGEYRNPEDIVPSNPTD 180
 QY 180 SGSSSEBSSTSGGY-IFYT-FSTVHPIDEDSPWTTDSIDRIPATNMDSSHSTLLQPTAN 237
 DB 181 SGSSSEBSSTSGGY-IFYT-FSTVHPIDEDSPWTTDSIDRIPATNMDSSHSTLLQPTAN 237
 QY 238 PNTGLVEDLDRGTPLSMTTQGSNSQSFSTSHGLEDKDPHTTSTLTSNRNDVYTGRRD 297
 DB 217 ----- 216

QY 298 PNHSEGTLLLEGTSYHPHTKSRTPVTSKTSFGVAVTVGDSNNVRSLSGQ 357
DB 217 --HEENT-----PTTKDQ-----GSDPH----- 233
QY 358 DTFPSGSHHTHGSSEDSHSGSQEGGANTSGPIRTPOIPEMILLIASLALAILAV 417
DB 234 -----SGRSHHTHGTESPGVSRSGQEGGANTSGPIRTPOIPEMILLIASLALAILAV 268
QY 418 CIANVSRRCGQKKCLVINGNAGVEDRKPSGLNGEASKSQEWHLVNNESSSTPDQFMT 477
DB 289 CIANVSRRCGQKKCLVINGNAGVEDRKASGLNGEASKSQEWHLVNNESSSTPDQFMT 348
QY 478 ADETNPLOVDMKIGV 493
DB 349 ADETNPLOVDMKIGV 364

RESULT 14

ID 092493 PRELIMINARY; PRT; 294 AA.
AC 092493;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Cell surface glycoprotein CD44.
GN CD44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RX MEDLINE=92017823; Pubmed=1922057;
RA Shivelman E, Bishop J M;
RT "Expression of CD44 is repressed in neuroblastoma cells."
RU Mol. Cell. Biol. 11:5446-5453(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RX MEDLINE=93101687; Pubmed=1465456;
RA Sreearon G.R., Bell M.V., Jackson D.G., Cornells F.B., Gerth U.,
RA Bell J.I.;
RT "Genomic structure of DNA encoding the lymphocyte homing receptor CD44
RT reveals at least 12 alternatively spliced exons."
RU Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164(1992).
DR EMBL; L05423; AAB13622.1; -.
DR EMBL; M69215; AAB13622.1; JOINED.
DR EMBL; L05407; AAB13622.1; JOINED.
DR EMBL; L05408; AAB13622.1; JOINED.
DR EMBL; L05409; AAB13622.1; JOINED.
DR EMBL; L05410; AAB13622.1; JOINED.
DR EMBL; L05420; AAB13622.1; JOINED.
DR EMBL; L05421; AAB13622.1; JOINED.
DR EMBL; L05422; AAB13622.1; JOINED.
DR PIR; JH0417; JH0417.
DR HSSP; P98066; ITSG.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001231; CD44_antigen.
DR InterPro; IPR000538; Link.
DR Pfam; PF00193; Xlink; 1.
DR PRINTS; PR00658; CD44.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00445; Link; 1.
DR PROSITE; PS01241; Link; 1.
SQ SEQUENCE 294 AA; 32074 MW; E11B9B38FA74817 CRC64;

Query Match 57.3%; Score 1494; DB 4; Length 294;
Best Local Similarity 68.7%; Pred. No. 2.8e-98;

Matches 292; Conservative 0; Mismatches 1; Indels 132; Gaps 1;
QY 1 MDKFWHAAMGICLVPLSLAIDINTCRFAGVHEVKNGRYSISRTAADLCAFNSTL 60
DB 1 MDKFWHAAMGICLVPLSLAIDINTCRFAGVHEVKNGRYSISRTAADLCAFNSTL 60
QY 61 PTMAQMERALSIGETCKYGFIBGVVYPRIRHPSICAAANTGVYILTSNTSOYDTYCFN 120
DB 61 PTMAQMERALSIGETCKYGFIBGVVYPRIRHPSICAAANTGVYILTSNTSOYDTYCFN 120
QY 121 ASAPPEEDCTSVTLLENAPDGPITITVNRDGTYYVQGEYRTNPEDIPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTLLENAPDGPITITVNRDGTYYVQGEYRTNPEDIPSNPTDDVSS 180
QY 181 GSSSERSTSGGYIFTFSTYVHPIDPDESPWITDSTORIPATNMDSHSTLLOFTANPNT 240
DB 181 GSSSERSTSGGYIFTFSTYVHPIDPDESPWITDSTORIPATNMDSHSTLLOFTANPNT 240
QY 241 GLVEDLDRGTPLSMTTQGSNSGFSSTSHGLEDKDHPTTSTLTSNRNDVTGGRDPNH 300
DB 224 -----DQDTF 228
QY 301 SEGSTTLLEGTSYHPHTKSRTPVTSKTSFGVAVTVGDSNNVRSLSGQDPTF 360
DB 229 HPSGSHHTHGSSEDSHSGSQEGGANTSGPIRTPOIPEMILLIASLALAILAVCIA 288
QY 421 VNRR 425
DB 289 VNRR 293

RESULT 15

ID 070509 PRELIMINARY; PRT; 364 AA.
AC 070509;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Glycoprotein CD44s (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Lumbar spine;
RA Stevens J.W.;
RT Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065147; AAC17117.1; -.
DR HSSP; P98066; ITSG.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001231; CD44_antigen.
DR InterPro; IPR000538; Link.
DR Pfam; PF00193; Xlink; 1.
DR PRINTS; PR00658; CD44.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00445; Link; 1.
DR PROSITE; PS01241; Link; 1.
SQ SEQUENCE 364 AA; 39725 MW; BA249776C4419AAT7 CRC64;

Query Match 52.4%; Score 1367.5; DB 11; Length 364;
Best Local Similarity 59.2%; Pred. No. 3.6e-89;
Matches 295; Conservative 20; Mismatches 44; Indels 139; Gaps 10;
QY 1 MDKFWHAAMG-ICLVPLSLA--QIDINTCRFAGVHEVKNGRYSISRTAADLCAFN 57


```
Db      1 MDKVMHTANGLLCLLOLSIAOOIDINTICRYAGVFEHKNRYSISRTAEADLCEAFN 60
Qy      58 STLPYMAQMEKALSIGETCRYGFIIEGHVVIPIRIHNSICAANTGVYI-LTSNTSQYDT 116
Db      61 TLPLTMAQMLASKGEFCRYGFIIEGHVVIPIRIHNAICAAANTGVYIILASNTSHYDT 120
Qy      117 YCFNAPAPPEEDCTSTDLFNAPDGPITITIVNRDGRIVQKGEYRINPEDIYPSNYPTD 176
Db      121 YCFNAPAPPEEDCTSTDLFNAPDGPITITIVNRDGRIVQKGEYRINPEDIYPSNYPTD 180
Qy      177 DVSSGSSRSSTSGYIFFTFSTVHPIDEDSPWITDSTDRIPATNMDSHSTTLQPTA 236
Db      181 DVSSGSTIEK-STIEGYILMT-----DLP-----TSQPT- 208
Qy      237 NPNTGLVEDLDRTGFLSMTTQGSNSQSFSSTHEGLEBDKDHPTTSTLSSNRNDVTGRR 296
Db      209 -----GDR 211
Qy      297 DPNHSEGSTTILLEGYTSHPHTYKSRFTIPVTSAKTGFQVTAVTAGDSNVNRSLSGD 356
Db      212 DDAFFIGST-----LATSDGSSMD----- 231
Qy      357 QDTFHPGGSHT-THGESDQSHSGQEGGANTSGPIRTPOIPEWLIILASILALALIL 415
Db      232 ----PRGFDVTWHSGLAGHSSGNQDSGVTTSQPARRPQIPEWLIILASILALALIL 286
Qy      416 AVCIAVNSRRRCQKKLVINSNGAVEDRKPSELNGEASKSOEWHLVNXESSETPDQF 475
Db      287 AVCIAVNSRRRCQKKLVINSNGAVEDRKPSELNGEASKSOEWHLVNXESSETPDQF 346
Qy      476 MTADETRNLOVDMKIGV 493
Db      347 MTADETRNLOVDMKIGV 364
```

Search completed: March 8, 2004, 06:11:38
Job time : 50 secs

CC disorder amenable for treatment with a stem cell in a subject or mammal,
 CC where the disorder is selected from myocardial infarction, Parkinson's
 CC disease, diabetes, congenital muscle dystrophy, stroke,
 CC genetic/congenital disorders and liver disorders. Another method
 CC disclosed in the invention is useful for increasing levels of engrafted
 CC stem cells in a subject, preferably a human suffering from or at risk for
 CC hematopoietic disorder or cancer, preferably blood cancer, respectively.
 CC The HCEIL protein is useful for treating hematopoietic disorders such as
 CC aplastic anaemia, and inflammatory disorders such as rheumatoid
 CC arthritis, inflammatory bowel disease and asthma. The HCEIL protein is
 CC useful as an immunogen for producing anti-HCEIL antibodies. The present
 CC sequence represents the HCEIL protein of the invention

XX Sequence 493 AA;

Query Match 100.0%; Score 2608; DB 5; Length 493;
 Best Local Similarity 100.0%; Pred. No. 1,2e-191;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKFWMAAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60
 DB 1 MDKFWMAAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60
 QY 61 PTMAQMEKALSIGETCYGFIETGVYVPIRHHNSICANNNGVYILTSNTSQDYTCFN 120
 DB 61 PTMAQMEKALSIGETCYGFIETGVYVPIRHHNSICANNNGVYILTSNTSQDYTCFN 120
 QY 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSS 180
 QY 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240
 DB 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240
 QY 241 GLVEDLDRTGFLSWTTOQSNQSFSTSHGLEDKCHPTTSTLSSNRNDVTCGRDPNH 300
 DB 241 GLVEDLDRTGFLSWTTOQSNQSFSTSHGLEDKCHPTTSTLSSNRNDVTCGRDPNH 300
 QY 301 SEGSTTLLEGTSYRHTKESRPTIPYTSKTSFGYTAATYVGDNSNNRSLSGQDPTF 360
 DB 301 SEGSTTLLEGTSYRHTKESRPTIPYTSKTSFGYTAATYVGDNSNNRSLSGQDPTF 360
 QY 361 HPSGSHTTGSESDGSHSQSGEGGANTSGPRTPOIPMLIILSLALAILAVCIA 420
 DB 361 HPSGSHTTGSESDGSHSQSGEGGANTSGPRTPOIPMLIILSLALAILAVCIA 420
 QY 421 VNSRRRCGQKKKLYVINGNVAVEDRKPSSGLNGEASKSQEVAHLVNRKSSSTPPQNTABE 480
 DB 421 VNSRRRCGQKKKLYVINGNVAVEDRKPSSGLNGEASKSQEVAHLVNRKSSSTPPQNTABE 480
 QY 481 TRNLQVDMKIGV 493
 DB 481 TRNLQVDMKIGV 493

RESULT 2
 ID ABU04623 standard; protein; 493 AA.
 AC ABU04623;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1289.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

XX WO200278524-A2.
 PN 10-OCT-2002.
 XX 28-MAR-2002; 2002WO-US009671.
 XX 28-MAR-2001; 2001US-0279495P.
 XX 21-MAY-2001; 2001US-0292544P.
 XX 08-AUG-2001; 2001US-0310801P.
 XX 01-OCT-2001; 2001US-0326370P.
 XX 04-DEC-2001; 2001US-0356780P.
 XX 20-FEB-2002; 2002US-0358985P.
 PA (ZYCO-) ZYCO5 INC.
 XX Chicx RM, Tomlinson AJ, Urban RG;
 XX WPI; 2003-040607/03.
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX Example 2; SEQ ID NO 1289; 134pp; English.

The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that bind to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SO Sequence 493 AA;

Query Match 100.0%; Score 2608; DB 6; Length 493;

Best Local Similarity 100.0%; Pred. No. 1,2e-191;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKFWMAAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60
 DB 1 MDKFWMAAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60

QY 61 PTMAQMEKALSIGETCYGFIETGVYVPIRHHNSICANNNGVYILTSNTSQDYTCFN 120
 DB 61 PTMAQMEKALSIGETCYGFIETGVYVPIRHHNSICANNNGVYILTSNTSQDYTCFN 120

QY 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTYYVQGEYRTNPEDIYPSNPTDDVSS 180

QY 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240
 DB 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240

QY 241 GLVEDLDRTGFLSWTTOQSNQSFSTSHGLEDKCHPTTSTLSSNRNDVTCGRDPNH 300
 DB 241 GLVEDLDRTGFLSWTTOQSNQSFSTSHGLEDKCHPTTSTLSSNRNDVTCGRDPNH 300
 QY 301 SEGSTTLLEGTSYRHTKESRPTIPYTSKTSFGYTAATYVGDNSNNRSLSGQDPTF 360
 DB 301 SEGSTTLLEGTSYRHTKESRPTIPYTSKTSFGYTAATYVGDNSNNRSLSGQDPTF 360

Db 301 SEGSTTLLEGYSHYPHRKESRTFIPVTSAKTSGFVTAIVYGDNSNVNRSLSGDDDTF 360
 QY 361 HPSGGSHTTGSSEDSHSGSQEGANTTSGPIRTPQIPEWIIIIASIIALALIIAVCIA 420
 Db 361 HPSGGSHTTGSSEDSHSGSQEGANTTSGPIRTPQIPEWIIIIASIIALALIIAVCIA 420
 QY 421 VNSRRRCGQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVNKESSETPDDQMTADE 480
 Db 421 VNSRRRCGQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVNKESSETPDDQMTADE 480
 QY 481 TRNLQNVDMKIGV 493
 Db 481 TRNLQNVDMKIGV 493
 RESULT 3
 AB004612
 ID AB004612 standard; protein; 493 AA.
 XX
 AC AB004612;
 XX
 DT 29-JAN-2003 (first entry)
 XX
 DE Human expressed protein tag (EPT) #1278.
 XX
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200278524-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US009671.
 XX
 PR 28-MAR-2001; 2001US-0279495P.
 XX
 PR 21-MAY-2001; 2001US-0292544P.
 XX
 PR 08-AUG-2001; 2001US-0310801P.
 XX
 PR 01-OCT-2001; 2001US-0326370P.
 XX
 PR 04-DEC-2001; 2001US-0336780P.
 XX
 PR 20-FEB-2002; 2002US-0358985P.
 XX
 PA (ZYCO-) ZYCOS INC.
 XX
 PI Chicx RM, Tomlinson AJ, Urban RG;
 XX
 DR WPI; 2003-040607/03.
 XX
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX
 PS Example 2; SEQ ID NO 1278; 134PD; English.
 CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide that binds to this
 CC polypeptide is useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIP0 at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 493 AA;
 Query Match 100.0%; Score 2608; DB 6; Length 493;
 Best Local Similarity 100.0%; Pred. No.1,2e-191;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDKFWHAAAGCLVPLSLAQIDNLITCSPAGVFHYEKGRYSISRTEADICKANSTL 60
 Db 1 MDKFWHAAAGCLVPLSLAQIDNLITCSPAGVFHYEKGRYSISRTEADICKANSTL 60
 QY 61 PTMAQMEKALSTIGFETCRYGFTIGHVIVIRIHNSICAAANTGVYILTNTSOYPTTCFN 120
 Db 61 PTMAQMEKALSTIGFETCRYGFTIGHVIVIRIHNSICAAANTGVYILTNTSOYPTTCFN 120
 QY 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
 Db 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
 QY 181 GSSSESSSTSGGYTFYTFSTVHPIDPDSPWINDSTDRIPATNMDSSHSTTLQPTANPMT 240
 Db 181 GSSSESSSTSGGYTFYTFSTVHPIDPDSPWINDSTDRIPATNMDSSHSTTLQPTANPMT 240
 QY 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEEKDHPPTSTLSSNRNDVGTGRDPN 300
 Db 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEEKDHPPTSTLSSNRNDVGTGRDPN 300
 QY 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEEKDHPPTSTLSSNRNDVGTGRDPN 300
 Db 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEEKDHPPTSTLSSNRNDVGTGRDPN 300
 QY 301 SEGSTTLLEGYSHYPHRKESRTFIPVTSAKTSGFVTAIVYGDNSNVNRSLSGDDDTF 360
 Db 301 SEGSTTLLEGYSHYPHRKESRTFIPVTSAKTSGFVTAIVYGDNSNVNRSLSGDDDTF 360
 QY 361 HPSGGSHTTGSSEDSHSGSQEGANTTSGPIRTPQIPEWIIIIASIIALALIIAVCIA 420
 Db 361 HPSGGSHTTGSSEDSHSGSQEGANTTSGPIRTPQIPEWIIIIASIIALALIIAVCIA 420
 QY 421 VNSRRRCGQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVNKESSETPDDQMTADE 480
 Db 421 VNSRRRCGQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVNKESSETPDDQMTADE 480
 QY 481 TRNLQNVDMKIGV 493
 Db 481 TRNLQNVDMKIGV 493
 RESULT 4
 AAR20817
 ID AAR20817 standard; protein; 493 AA.
 XX
 AC AAR20817;
 XX
 DT 25-MAR-2003 (revised)
 DT 21-MAY-1992 (first entry)
 XX
 DE Epithelial CD44 Antigen.
 XX
 KW Rapid immunoselection cloning technique; cell surface antigen;
 KW epithelium; tumour invasiveness; carcinoma.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Peptide 1..20
 XX Domain /label= signal
 XX Modified-site 21..400
 XX /label= extracellular
 XX Modified-site 65..67
 XX /label= N-linked glycosylation
 XX /note= "putative"
 XX Modified-site 97..99
 XX /label= N-linked glycosylation

```

FT      /note= "putative"
FT      140..142
FT      Modified-site
FT      /label= "N-linked glycosylation"
FT      /note= "putative"
FT      150..152
FT      Modified-site
FT      /label= "N-linked glycosylation"
FT      /note= "putative"
FT      160..162
FT      Modified-site
FT      /label= "N-linked glycosylation"
FT      /note= "putative"
FT      180..181
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      190..191
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      295..297
FT      Modified-site
FT      /label= "N-linked glycosylation"
FT      /note= "putative"
FT      390..391
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      401..421
FT      Region
FT      /label= transmembrane
FT      422..493
FT      Domain
FT      /label= cytoplasmic

PN      W09201049-A.
XX      23-JAN-1992.
XX      13-JUL-1990; 90US-00553759.
XX      13-JUL-1990; 90US-00553759.
XX      13-JUL-1990; 90US-00553759.
XX      (GENO ) GEN HOSPITAL CORP.
XX      Seed B, Aruffo A, Amiot M;
XX      WPI, 1992-056864/07.
XX      N-PSDB; AAQ21186.
XX      New CD53 cell surface antigen and DNA encoding it - for immuno-therapy
XX      and diagnosis of haematopoietic neoplasms, etc.
XX      Example 15; Page 117; 160pp; English.
XX      The cDNA encoding epithelial CD44 is quite similar to haematopoietic
XX      CD44.5 cDNA (see AAQ21185), but encodes an additional extracellular
XX      domain of 165 amino acids, inserted about 140 residues upstream of the
XX      transmembrane section shared by both clones. The ability to interfere
XX      with binding of epithelial CD44 to extracellular matrices can be useful
XX      in diagnostics and therapy. e.g. to diminish the likelihood of metastasis
XX      in cancer patients. Soluble forms of CD44 can act to prevent metastatic
XX      cells from "homing" to lymph nodes. (Updated on 25-MAR-2003 to correct PA
XX      field.)
XX      Sequence 493 AA;

```

```

Query Match          99.3%; Score 2590; DB 2; Length 493;
Best Local Similarity 99.4%; Pred. No. 2,9e-150;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 MDKFWMAWGLCVPLSLAQLDINTCFAGVHFVKNRGYSISRTAADLCAFNSTL 60
DB      1 MDKFWMAWGLCVPLSLAQLDINTCFAGVHFVKNRGYSISRTAADLCAFNSTL 60
QY      61 PTMAQMEKALISIGFTCRYGFTBGHVVPRIHPSICAAANTGVYIITSNTSOYDTYCFN 120
DB      61 PTMAQMEKALISIGFTCRYGFTBGHVVPRIHPSICAAANTGVYIITSNTSOYDTYCFN 120

```

```

QY      121 ASAPPEEDCTSVTDLPNAPDGPITITTVNRDGRYVKGGRYNPEDITYSNNPTDDVSS 180
DB      121 ASAPPEEDCTSVTDLPNAPDGPITITTVNRDGRYVKGGRYNPEDITYSNNPTDDVSS 180
QY      181 GSSSESRSSSTGGYVFPYESTVHPIDPDSPMITPSTDRIPATNNDSSHSTTLOPTANPT 240
DB      181 GSSSESRSSSTGGYVFPYESTVHPIDPDSPMITPSTDRIPATNNDSSHSTTLOPTANPT 240
QY      241 GLVEDLDRTGPLSMITTOQSNQSFSSTSGHEGLEDKDPHTSTLTSSRNNDYTGGRDPNH 300
DB      241 GLVEDLDRTGPLSMITTOQSNQSFSSTSGHEGLEDKDPHTSTLTSSRNNDYTGGRDPNH 300
QY      301 SEGSTTLLEGTSYHPHTKESRTIPYTSATKSGFYTAAYVGDPSNVNRSLSGDDPTF 360
DB      301 SEGSTTLLEGTSYHPHTKESRTIPYTSATKSGFYTAAYVGDPSNVNRSLSGDDPTF 360
QY      361 HPSGGSHHTSGESDGHSGSQEGANTTSGPIRTPOIPEWLTILASILALAILAVCIA 420
DB      361 HPSGGSHHTSGESDGHSGSQEGANTTSGPIRTPOIPEWLTILASILALAILAVCIA 420
QY      421 VNSRRRCQKKLVVNSGNGAVEDRKSGGNGEASKCEMVHLVNXKSESTPDQFMTADE 480
DB      421 VNSRRRCQKKLVVNSGNGAVEDRKSGGNGEASKCEMVHLVNXKSESTPDQFMTADE 480
QY      481 TRNLQNVDMKIGV 493
DB      481 TRNLQNVDMKIGV 493

```

```

RESULT 5
AAW80454
AAW80454 standard; protein; 493 AA.
XX
XX      AAW80454;
AC      25-MAR-2003 (revised)
DT      07-JUN-1999 (first entry)
XX
XX      Human CD44 antigen (epithelial form).
XX      CD44; cell surface antigen; human; cDNA library; T lymphocyte;
XX      metastasis; therapy.
XX      Homo sapiens.
XX
XX      Key Location/Qualifiers
XX      FT 401..421
XX      FT Domain /note= "transmembrane domain"
XX
XX      US5830731-A.
XX
XX      03-NOV-1998.
XX
XX      21-MAY-1997; 97US-00861205.
XX
XX      25-FEB-1988; 88US-00160416.
XX      13-JUL-1989; 89US-00379076.
XX      23-MAR-1990; 90US-00498809.
XX      13-JUL-1990; 90US-00553759.
XX      01-DEC-1992; 92US-00983647.
XX
XX      (GENO ) GEN HOSPITAL CORP.
XX      Seed B, Aruffo A;
XX
XX      WPI, 1998-609251/51.
XX      N-PSDB; AAV63462.
XX
XX      New cloning vector and polylinker - based on existing sequences for
XX      efficient cloning and expression of mammalian cDNA(s), especially human
XX      lymphocyte antigenic sequences.

```

PS Example 15; Col 73-76; 75pp; English.

CC This polypeptide comprises human CD4 antigen. Its amino acid sequence
CC was deduced from the nucleotide sequence (see AAV63462) of a cDNA clone
CC isolated from a colon carcinoma HT29 cell library using a novel method
CC for cloning cDNAs from mammalian expression libraries. The method is
CC based on transient expression of an antigen in eukaryotic cells and
CC physical selection of cells expressing the antigen by adhesion to an
CC antibody-coated substrate. It is useful for the isolation and cloning of
CC any protein which can be expressed and transported to the cell surface
CC membrane of a eukaryotic cell, and was used to clone genes (see AAV63442-
CC 63) encoding cell surface antigens from mammalian lymphocytes (see
CC AAV60440-55). The purified genes and proteins are useful for
CC immunodiagnostic and immunotherapeutic applications, including the
CC diagnosis and treatment of immune-mediated infections, diseases, and
CC disorders of animals, including humans. The epithelial form of CD4 is a
CC glycosylated protein of about 160 kDa. The ability to interfere with the
CC binding of epithelial CD4 with extracellular matrices may be useful in
CC therapy or diagnosis, e.g. to diminish the likelihood of metastasis in
CC cancer patients. Soluble forms of CD4 can act to prevent metastatic
CC cells from homing to lymph nodes. (Updated on 23-MAR-2003 to correct PR
CC field.)

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 2; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190; Indels 0; Gaps 0;

Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWHAAAGLCVPLSLAIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
DB 1 MDKFWHAAAGLCVPLSLAIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
QY 61 PTMAQMEKALSIGETCRVGFIEGHVVIPIRHNSICANNNGVYILNTSQTGYTCFN 120
DB 61 PTMAQMEKALSIGETCRVGFIEGHVVIPIRHNSICANNNGVYILNTSQTGYTCFN 120
QY 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180
QY 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180
QY 181 GSSSERSTSGGYIFTFSTVHPPIEDSDPWITDSIDRIATNMDSHSTLQPTANPMT 240
DB 181 GSSSERSTSGGYIFTFSTVHPPIEDSDPWITDSIDRIATNMDSHSTLQPTANPMT 240
QY 241 GLVEDLDRGTPLSMITTOQNSQSFSTSHGLEEDKHPITSTLTSSNRNDVTCGRADPNH 300
DB 241 GLVEDLDRGTPLSMITTOQNSQSFSTSHGLEEDKHPITSTLTSSNRNDVTCGRADPNH 300
QY 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFVTAIVGDSNNVNRSLSGQDTF 360
DB 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFVTAIVGDSNNVNRSLSGQDTF 360
QY 361 HPSSGSHTTGSESDGSHSGOEGAGNTSGPIRTPQIPMLIILSLALATILAVCIA 420
DB 361 HPSSGSHTTGSESDGSHSGOEGAGNTSGPIRTPQIPMLIILSLALATILAVCIA 420
QY 421 VNSRRRCGQKKKLVINSNGAVEDRKPSSGLNGEASKSQEVAHLVNKESSTPQFWTADE 480
DB 421 VNSRRRCGQKKKLVINSNGAVEDRKPSSGLNGEASKSQEVAHLVNKESSTPQFWTADE 480
QY 481 TRNLQNDPMKIGV 493
DB 481 TRNLQNDPMKIGV 493

RESULT 6
AAW89151 ID AAW89151 standard; protein; 493 AA.

XX AAW89151;
XX 10-MAY-1999 (first entry)

XX Human CD44 antigen (epithelial form).
DE CD44; cell surface antigen; human; cDNA library; T lymphocyte;
KW metastasis; therapy.

XX Homo sapiens.

XX Key Location/Qualifiers
XX Domain 401..421
XX /note="transmembrane domain"

XX US5849898-A.

XX 15-DEC-1998.

XX 07-JUN-1995; 95US-00485447.

XX 25-FEB-1988; 68US-00160416.

XX 13-JUN-1989; 89US-00379076.

XX 23-MAR-1990; 90US-00488809.

XX 13-JUN-1990; 90US-00553759.

XX 01-DEC-1992; 92US-00983647.

XX (GENO) GEN HOSPITAL CORP.

XX Seed B, Oquendo C, Camerini D, Stramenkovic I, Stengelin S;

XX Amiot M, Lauffer L, Allen J, Simons D, Aruffo A;

XX WPI; 1999-069813/06.

XX N-PEDB; AAV81219.

XX cDNA encoding human CD40 antigen - useful for cloning cDNA encoding cell

XX surface antigens, constructing cDNA libraries, expressing vectors for

XX expression in eukaryotic cells or their fragments.

XX Example 15; Col 79-82; 79pp; English.

CC This polypeptide comprises human CD4 antigen. Its amino acid sequence
CC was deduced from the nucleotide sequence (see AAV81219) of a cDNA clone
CC isolated from a colon carcinoma HT29 cell library using a novel method
CC for cloning cDNAs from mammalian expression libraries. The method is
CC based on transient expression of an antigen in eukaryotic cells and
CC physical selection of cells expressing the antigen by adhesion to an
CC antibody-coated substrate. It is useful for the isolation and cloning of
CC any protein which can be expressed and transported to the cell surface
CC membrane of a eukaryotic cell, and was used to clone genes (see AAV81198-
CC 220) encoding cell surface antigens such as CD1a, CD1b, CD1c, CD2, CD6,
CC CD7, CD13, CD14, CD16, CD19, CD20, CD22, CD26, CD27, CD28, CD31, CD32a,
CC CD32b, CD33, CD34, CD36, CD37, CD38, CD39, CD40, CD43, CD44, CD53, ICAM,
CC LFA-3, FCRII, FCRIIb, TILSA and Leu8 (see AAW81198) is specifically claimed. The
CC AAW88451). CD40 cDNA (see AAW81198) is specifically claimed. The
CC epithelial form of CD44 is a glycosylated protein of about 160 kDa. The
CC ability to interfere with the binding of epithelial CD44 with
CC extracellular matrices may be useful in therapy or diagnosis, e.g. to
CC diminish the likelihood of metastasis in cancer patients. Soluble forms
CC of CD44 can act to prevent metastatic cells from homing to lymph nodes

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 2; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190; Indels 0; Gaps 0;

Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWHAAAGLCVPLSLAIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
DB 1 MDKFWHAAAGLCVPLSLAIDNITCRFAGVHVENKGRYSISRTEADLCRAFNSTL 60
QY 61 PTMAQMEKALSIGETCRVGFIEGHVVIPIRHNSICANNNGVYILNTSQTGYTCFN 120
DB 61 PTMAQMEKALSIGETCRVGFIEGHVVIPIRHNSICANNNGVYILNTSQTGYTCFN 120
QY 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180

```

Db 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPDPIDVPSNPTDDVSS 180
Qy 181 GSSSERSTSGGYFYFSTVHP1PDEDSPMITDSTDR1PATNDSHSTTLOPTANPNT 240
Db 181 GSSSERSTSGGYFYFSTVHP1PDEDSPMITDSTDR1PATNDSHSTTLOPTANPNT 240
Qy 241 GVEDLDRTGPIPLSMTTQSNQSPFSTHGELEEDKHPPTSTLTSSNRNDVTGRRDPNH 300
Db 241 GVEDLDRTGPIPLSMTTQSNQSPFSTHGELEEDKHPPTSTLTSSNRNDVTGRRDPNH 300
Qy 301 SEGSTTLLEGYTSHPHTKESRRTFIPVTSKATGSPGYTAVTVGDSNNVRSLSGDDDTF 360
Db 301 SEGSTTLLEGYTSHPHTKESRRTFIPVTSKATGSPGYTAVTVGDSNNVRSLSGDDDTF 360
Qy 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPPEW11LALAL1LAVCIA 420
Db 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPPEW11LALAL1LAVCIA 420
Qy 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNRKESSTPDQFTADE 480
Db 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNRKESSTPDQFTADE 480
Qy 481 TRNLQNVDMKIGV 493
Db 481 TRNLQNVDMKIGV 493

RESULT 7
AAV96140 standard; protein; 493 AA.
ID AAV96140;
AC AAV96140;
XX
XX 19-DEC-2000 (first entry)
DT
XX Human epithelial CD44.
DE
XX CD44.5; cell surface antigen; human; immunoselection; panning;
KW immunodiagnosis; diagnosis; immunotherapy; gene therapy; immune disorder;
KW infection; asthma; immune-complex disease; amyloidosis;
KW multiple sclerosis; parasitic disease; epithelial; metastasis.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1.19
FT Peptide /label= signal_peptide
FT 20.493
FT Protein /label= Mature_protein
FT 401.421
FT Domain /label= Transmembrane_domain
FN US611093-A.
XX
XX 29-AUG-2000.
PD
XX 28-OCT-1998; 98US-00181612.
PF
XX 25-FEB-1988; 88US-00160416.
PR 13-JUL-1989; 89US-00379076.
PR 23-MAR-1990; 90US-00498809.
PR 13-JUL-1990; 90US-00537359.
PR 01-DEC-1992; 92US-00983647.
XX
XX (GCHO) GEN HOSPITAL CORP.
PA
XX Stamenkovic I, Seed B;
PI
XX WPI; 2000-586382/55.
DR N-PSDB; AAA50596.
XX
XX Isolated nucleic acid molecule encoding the CD19 cell surface antigen,

```

```

PT useful for immunodiagnosis and immunotherapy of immune-mediated
PT infections or disorders, e.g. asthma, immune-complex disease, parasitic
PR diseases.
PS Example 15; Col 75-78; 75pp; English.
XX
XX The present sequence is that of human epithelial CD44, a glycosylated
CC membrane protein that may promote tumor invasiveness. The sequence was
CC predicted from a cDNA clone (see AAA50598) isolated by screening a cDNA
CC library prepared from the the colon carcinoma line HT29 using the method
CC of the invention. The method, designed to isolate cell surface antigen
CC (CSA) nucleic acids, is based upon transient expression of a CSA in
CC eukaryotic cells and physical selection of cells expressing the antigen
CC by adhesion to (panning on) an antibody-coated substrate such as a
CC culture dish. CSA nucleic acids isolated by the method of the invention,
CC and the proteins they encode, are useful for immunodiagnosis and
CC immunotherapeutic applications, including the diagnosis and treatment of
CC immune-mediated infections, diseases, and disorders in animals, including
CC humans. These disorders include asthma, immune-complex disease,
CC amyloidosis, parasitic diseases or multiple sclerosis. The ability to
CC interfere with the binding of epithelial CD44 with extracellular matrices
CC can be useful in diagnostics and therapy. Interference of the binding can
CC diminish the likelihood of metastasis in cancer patients. Soluble forms
CC of CD44 can act to prevent metastatic cells from homing to lymph nodes
XX
XX Sequence 493 AA;
SQ
Query Match 99.3%; Score 2590; DB 3; Length 493;
Best Local Similarity 99.4%; Pred. No. 2.9e-190;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MDKFWHAAWGLCVPLSLAQIDLNTCRFAGVHVHKNRYSISRTAADLCAENSTL 60
Db 1 MDKFWHAAWGLCVPLSLAQIDLNTCRFAGVHVHKNRYSISRTAADLCAENSTL 60
Qy 61 PTMAOMEXALSGEFTGRCYRIGBHVVIPRIHPSICANNVGYLLTSTSOVDTCEN 120
Db 61 PTMAOMEXALSGEFTGRCYRIGBHVVIPRIHPSICANNVGYLLTSTSOVDTCEN 120
Qy 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPDPIDVSS 180
Db 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPDPIDVSS 180
Qy 181 GSSSERSTSGGYFYFSTVHP1PDEDSPMITDSTDR1PATNDSHSTTLOPTANPNT 240
Db 181 GSSSERSTSGGYFYFSTVHP1PDEDSPMITDSTDR1PATNDSHSTTLOPTANPNT 240
Qy 241 GVEDLDRTGPIPLSMTTQSNQSPFSTHGELEEDKHPPTSTLTSSNRNDVTGRRDPNH 300
Db 241 GVEDLDRTGPIPLSMTTQSNQSPFSTHGELEEDKHPPTSTLTSSNRNDVTGRRDPNH 300
Qy 301 SEGSTTLLEGYTSHPHTKESRRTFIPVTSKATGSPGYTAVTVGDSNNVRSLSGDDDTF 360
Db 301 SEGSTTLLEGYTSHPHTKESRRTFIPVTSKATGSPGYTAVTVGDSNNVRSLSGDDDTF 360
Qy 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPPEW11LALAL1LAVCIA 420
Db 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPPEW11LALAL1LAVCIA 420
Qy 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNRKESSTPDQFTADE 480
Db 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNRKESSTPDQFTADE 480
Qy 481 TRNLQNVDMKIGV 493
Db 481 TRNLQNVDMKIGV 493

RESULT 8
AAU02449 standard; protein; 493 AA.
ID AAU02449
XX
XX AAU02449;
AC

```

XX		29-AUG-2001 (first entry)
DT		
XX	DE	Human epithelial antigen CD44 polypeptide.
KW		Human, epithelial antigen; immune-mediated disease; CD44; infection;
KW		immune deficiency disorder; hypersensitivity; inflammation;
KW		systemic lupus erythematosus; platelet disorder; rheumatoid arthritis;
KW		transplant rejection; asthma.
XX		
OS		Homo sapiens.
XX		
FH		
FH	Key	Location/Qualifiers
FT	Domain	1..400
FT		/label= Extracellular_domain
FT	Modified-site	65..67
FT		/note= "N-linked glycosylation site"
FT	Modified-site	97..99
FT		/note= "N-linked glycosylation site"
FT	Modified-site	140..142
FT		/note= "N-linked glycosylation site"
FT	Modified-site	150..152
FT		/note= "N-linked glycosylation site"
FT	Modified-site	160..162
FT		/note= "N-linked glycosylation site"
FT	Binding-site	180..181
FT		/note= "SG dipeptide that forms the minimal attachment
FT		site of serine-linked chondroitin sulphate in
FT	Binding-site	proteoglycan proteins"
FT		190..191
FT		/note= "SG dipeptide that forms the minimal attachment
FT		site of serine-linked chondroitin sulphate in
FT	Binding-site	proteoglycan proteins"
FT		390..391
FT		/note= "SG dipeptide that forms the minimal attachment
FT		site of serine-linked chondroitin sulphate in
FT	Domain	proteoglycan proteins"
FT		401..421
FT		/label= Transmembrane_domain
FT	Modified-site	427..429
FT		/note= "N-linked glycosylation site"
XX		
PN		US6218525-B1.
XX		
PD		17-APR-2001.
XX		
PF		01-DEC-1992; 92US-00983647.
XX		
PR		25-FEB-1988; 88US-00160416.
PR		13-JUN-1989; 89US-00379076.
PR		13-JUL-1990; 90US-00553755.
XX		
PA		(GENO) GEN HOSPITAL CORP.
XX		
E1		Seed B, Aruffo A, Simmons D;
XX		
DR		WPI; 2001-289848/30.
DR		N-PADB; AAS03192.
XX		
PT		New recombinant DNA encoding CD28 useful for diagnosing and treating
PT		immune-mediated diseases, infections or disorders, e.g. systemic lupus
PT		erythematosus, asthma, transplant rejection, rheumatoid arthritis.
XX		
PS		Example 15; Col 69-72; 72pp; English.
CC		The present sequence encoding for human epithelial antigen CD44 is 1 of
CC		various human lymphocyte cell surface antigen polypeptide sequences
CC		(AAU02433-AAU02452) described in the present invention. The invention
CC		relates to a novel method of cloning cDNA encoding cell surface antigen
CC		and efficient construction of cDNA libraries. Also described are 2
CC		expression vectors (AAS03171, AAS03174) which provide high level
CC		expression in eukaryotic host cells. A genetically engineered cDNA
CC		sequence encoding the CD28 amino acid extracellular domain sequence

Query	Match	99.3%	Score 2590;	DB 4;	Length 493;
Best Local Similarity	99.4%;	Pred. No. 2.9e-190;			
Matches 499;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0	
QY	1 MDKFTNHWANGCLVPLISLAQIDNITCRFAGVYHVKKNRYSISTREADLCKANSTL	60			
DB	1 MDKFTNHWANGCLVPLISLAQIDNITCRFAGVYHVKKNRYSISTREADLCKANSTL	60			
QY	61 PTMAQWEKALSIGFETCRVGFIEGHVYLPRIHPSISCAANNIGVYLITSNTSQDYTCFN	120			
DB	61 PTMAQWEKALSIGFETCRVGFIEGHVYLPRIHPSISCAANNIGVYLITSNTSQDYTCFN	120			
QY	121 ASAPREDCTSTYDLPNAPFDGPITITITVNRDGRYVOKGKRYNPEDIYPSNFTDDVSS	180			
DB	121 ASAPREDCTSTYDLPNAPFDGPITITITVNRDGRYVOKGKRYNPEDIYPSNFTDDVSS	180			
QY	181 GSSSESSSTSGGYIFETFSVTHPIPDDESPWITDSTDRIPATMDSHSITLQPTANPT	240			
DB	181 GSSSESSSTSGGYIFETFSVTHPIPDDESPWITDSTDRIPATMDSHSITLQPTANPT	240			
QY	241 GLVEDLDRGFLSMITTOOSNQSFSTHSEGLIEBKHPITTSILTSSNRDLYGGRDPNH	300			
DB	241 GLVEDLDRGFLSMITTOOSNQSFSTHSEGLIEBKHPITTSILTSSNRDLYGGRDPNH	300			
QY	301 SEGSTTLLEGYSHYHPHTKESRTEFLPYTSATGSGFVYATVVDSDNSNVRLISGDDTF	360			
DB	301 SEGSTTLLEGYSHYHPHTKESRTEFLPYTSATGSGFVYATVVDSDNSNVRLISGDDTF	360			
QY	361 HPGGSHHTHGESGSHSGHSGOEGCANTTSGPITPPIPMILITLALALAILIAYCTA	420			
DB	361 HPGGSHHTHGESGSHSGHSGOEGCANTTSGPITPPIPMILITLALALAILIAYCTA	420			
QY	421 VNSRRRCGKKCLVINSNGANGVEDRKPDSGLNGEASKSQEVMHLVNKESSETPDOFTADE	480			
DB	421 VNSRRRCGKKCLVINSNGANGVEDRKPDSGLNGEASKSQEVMHLVNKESSETPDOFTADE	480			
QY	481 TRTLQVNMKIGV	493			
DB	481 TRTLQVNMKIGV	493			
RESULT 9					
ID	ABU04637				
XX	ABU04637 standard; protein; 493 AA.				
XX	ABU04637;				
XX	29-JAN-2003 (first entry)				
DE	Human expressed protein tag (EPT) #1303.				
KM	Translational profiling; expressed protein tag; EPT; kinase; phosphatase;				
KM	protease; protease inhibitor; transporter; cytoskeletal protein;				
KM	receptor; transcription factor; cancer; MHC;				
KM	major histocompatibility complex; myeloma; colon cancer; gastric cancer;				
XX	adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.				
XX	Homo sapiens.				

XX MO200278524-A2.
 PN 10-OCT-2002.
 XX 28-MAR-2002; 2002WO-US009671.
 XX 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX (ZYCO-) ZYCOS INC.
 PA Chicx RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.
 DR
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX
 PS Example 2; SEQ ID NO 1303; 134pp; English.
 XX
 CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC myeloma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MDRKFWHAAWGLCLVPLSLAQIDINTCRFAGVHYVKNRGRYISRTAADLCKAFNSTL 60
 DB 1 MDRKFWHAAWGLCLVPLSLAQIDINTCRFAGVHYVKNRGRYISRTAADLCKAFNSTL 60
 QY PTMAWMEKALSIGETRCYRGPIEGHVIPIPHNSICANNNGVILTSWTSQYDYYCEN 120
 DB 61 PTMAWMEKALSIGETRCYRGPIEGHVIPIPHNSICANNNGVILTSWTSQYDYYCEN 120
 QY 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGRYVKGERTNPEDIYSPNPTDDVSS 180
 DB 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGRYVKGERTNPEDIYSPNPTDDVSS 180
 QY 181 GSSSERSTSGCYFYFESTYHPIPEDSPITDSTRIRATNMDSHSTTLQPTANPT 240
 DB 181 GSSSERSTSGCYFYFESTYHPIPEDSPITDSTRIRATNMDSHSTTLQPTANPT 240
 QY 241 GLVEDLDRGTPLSMITTOOSNSQSFSTSHGLIEDKXHPPTSTLTSNRNDVYTGRRDPNH 300
 DB 241 GLVEDLDRGTPLSMITTOOSNSQSFSTSHGLIEDKXHPPTSTLTSNRNDVYTGRRDPNH 300
 QY 301 SEGSTTLLEGYTSYHPTKESRTIPIYTSKATGSPGYTAVTVGDSNVNRSLSGQDPF 360
 DB 301 SEGSTTLLEGYTSYHPTKESRTIPIYTSKATGSPGYTAVTVGDSNVNRSLSGQDPF 360

DB 301 SEGSTTLLEGYTSYHPTKESRTIPIYTSKATGSPGYTAVTVGDSNVNRSLSGQDPF 360
 QY 361 HPSCGSHHTTGSESDGSHSGOEGCANTTSGPIRTPQIPEMILLASLALAILAVCIA 420
 DB 361 HPSCGSHHTTGSESDGSHSGOEGCANTTSGPIRTPQIPEMILLASLALAILAVCIA 420
 QY 421 VNSRRRCGCKKVIYNSGNANVDRKPSGLNGASAKSQEWVHLWNKSSSTPCQFMTADE 480
 DB 421 VNSRRRCGCKKVIYNSGNANVDRKPSGLNGASAKSQEWVHLWNKSSSTPCQFMTADE 480
 QY 481 TRNLQNDVMDKIGV 493
 DB 481 TRNLQNDVMDKIGV 493
 RESULT 10
 ABU04627
 ID ABU04627 standard; protein; 493 AA.
 XX
 AC ABU04627;
 XX
 DT 29-JAN-2003 (first entry)
 XX
 DE Human expressed protein tag (EPT) #1293.
 XX
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX
 OS Homo sapiens.
 PN WO200278524-A2.
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US009671.
 XX
 PR 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX
 PA (ZYCO-) ZYCOS INC.
 XX
 PI Chicx RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.
 DR
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX
 PS Example 2; SEQ ID NO 1293; 134pp; English.
 XX
 CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC myeloma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWMAAWGICLVPSLAQIDINTCRFAGVHYEKNGRYSISREADLCRAFNSTL 60
 DB 1 MDKFWMAAWGICLVPSLAQIDINTCRFAGVHYEKNGRYSISREADLCRAFNSTL 60
 QY 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHNSICANNNGVYILTSNTSOYDTYCFN 120
 DB 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHNSICANNNGVYILTSNTSOYDTYCFN 120
 QY 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGTFRVQKGYRTNPEDIYPSNPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGTFRVQKGYRTNPEDIYPSNPTDDVSS 180
 QY 181 GSSSERSTSGGYFYFTSTVHPIDPDSFWITDSTRIPATNDSHSTLLQPTANPMT 240
 DB 181 GSSSERSTSGGYFYFTSTVHPIDPDSFWITDSTRIPATNDSHSTLLQPTANPMT 240
 QY 241 GLVEDDLRTGPIPLSMWTTQSNQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300
 DB 241 GLVEDDLRTGPIPLSMWTTQSNQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300
 QY 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGYTAIVGDSNSVNRSLSGDDTF 360
 DB 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGYTAIVGDSNSVNRSLSGDDTF 360
 QY 361 HPSGGSHTTGSSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIILASIALALIIAVCIA 420
 DB 361 HPSGGSHTTGSSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIILASIALALIIAVCIA 420
 QY 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLYNKESSTPPDQFMTADE 480
 DB 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLYNKESSTPPDQFMTADE 480
 QY 481 TRNLQNVDMKIGV 493
 DB 481 TRNLQNVDMKIGV 493

RESULT 11
 AB004639 standard; protein; 493 AA.
 AC AB004639;
 XX 29-JAN-2003 (first entry)
 DB Human expressed protein tag (EPT) #1305.
 XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
 XX Homo sapiens.
 OS WO200278524-A2.
 PN 10-OCT-2002.
 PD 28-MAR-2002; 2002MO-US009671.
 XX 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR

PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCOs INC.

XX Chicx RM, Tomlinson AJ, Urban RG;

XX WPI, 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.

XX Example 2; SEQ ID NO 1305; 134p; English.

CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I for class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWMAAWGICLVPSLAQIDINTCRFAGVHYEKNGRYSISREADLCRAFNSTL 60
 DB 1 MDKFWMAAWGICLVPSLAQIDINTCRFAGVHYEKNGRYSISREADLCRAFNSTL 60
 QY 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHNSICANNNGVYILTSNTSOYDTYCFN 120
 DB 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHNSICANNNGVYILTSNTSOYDTYCFN 120
 QY 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGTFRVQKGYRTNPEDIYPSNPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGTFRVQKGYRTNPEDIYPSNPTDDVSS 180
 QY 181 GSSSERSTSGGYFYFTSTVHPIDPDSFWITDSTRIPATNDSHSTLLQPTANPMT 240
 DB 181 GSSSERSTSGGYFYFTSTVHPIDPDSFWITDSTRIPATNDSHSTLLQPTANPMT 240
 QY 241 GLVEDDLRTGPIPLSMWTTQSNQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300
 DB 241 GLVEDDLRTGPIPLSMWTTQSNQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300
 QY 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGYTAIVGDSNSVNRSLSGDDTF 360
 DB 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGYTAIVGDSNSVNRSLSGDDTF 360
 QY 361 HPSGGSHTTGSSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIILASIALALIIAVCIA 420
 DB 361 HPSGGSHTTGSSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIILASIALALIIAVCIA 420
 QY 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLYNKESSTPPDQFMTADE 480
 DB 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLYNKESSTPPDQFMTADE 480

QY 481 TRNLQNVDMKIGV 493
DB 481 TRNLQNVDMKIGV 493

RESULT 12

ABU04633 standard; protein; 493 AA.

AC ABU04633;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1299.

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

KM protease; protease inhibitor; transporter; cytoskeletal protein;

KM receptor; transcription factor; cancer; MHC;

KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;

KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

PN WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCOS INC.

PI Chicx RM, Tomlinson AJ, Urban RG;

PT WPI; 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
cytoskeletal proteins, receptors or transcription factors), useful for
treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
leukemia.

PS Example 2; SEQ ID NO 1299; 134pp; English.

CC The invention describes a purified polypeptide, which comprises a
fragment of a kinase, phosphatase, protease, protease inhibitor,
transporter, cytoskeletal protein, receptor or transcription factor. The
polypeptide is useful as an immunogenic composition for eliciting in a
mammal an immunogenic response directed against any of the purified

CC polypeptide. The purified polypeptide, or the antibody that binds to this
polypeptide, is useful for treating cancer. The polypeptide binds also
useful for identifying compounds that binds to a naturally processed

CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing

CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an
expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

QY 1 MDKFWMAAMGLCLVPLSLAQIDINTCFRAGVHVEKNGKYSISRTAADLCAFNSTL 60
DB 1 MDKFWMAAMGLCLVPLSLAQIDINTCFRAGVHVEKNGKYSISRTAADLCAFNSTL 60

QY 61 PTMAQMEKALISGFETCYGFIKGVHVIPIRHPNSICANNNTGYIILTSNSQYDYCFN 120

DB 61 PTMAQMEKALISGFETCYGFIKGVHVIPIRHPNSICANNNTGYIILTSNSQYDYCFN 120

QY 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGTYYVQKGYRTNPEDITYSPNPTDDVSS 180

DB 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGTYYVQKGYRTNPEDITYSPNPTDDVSS 180

QY 181 GSSSERSSSTGGYIPTYSTVHPIDPDESPWITSTDRIPATNDSHSTLOFTAMPNT 240

DB 181 GSSSERSSSTGGYIPTYSTVHPIDPDESPWITSTDRIPATNDSHSTLOFTAMPNT 240

QY 241 GLVEDLDRTGPLSMVTQGSNQSFSSTHGLLEDDHPTSTLTSSNNDVTGGRDPNH 300

DB 241 GLVEDLDRTGPLSMVTQGSNQSFSSTHGLLEDDHPTSTLTSSNNDVTGGRDPNH 300

QY 301 SEGSTTLLEGYSHYPRHKEARTPIPTSAKTGSGVTAIVVGSNSVNSLSGDDQTF 360

DB 301 SEGSTTLLEGYSHYPRHKEARTPIPTSAKTGSGVTAIVVGSNSVNSLSGDDQTF 360

QY 361 HPGGSHHTHGSSEDSHSGSOGGANTSGPIRTPOIPEWLIILASLALALIAVCIA 420

DB 361 HPGGSHHTHGSSEDSHSGSOGGANTSGPIRTPOIPEWLIILASLALALIAVCIA 420

QY 421 VNSRRRCQCKKLIVNSGNGAVEDKPKGNGEAKSOMVHVNKSESPTDPQMTDE 480

DB 421 VNSRRRCQCKKLIVNSGNGAVEDKPKGNGEAKSOMVHVNKSESPTDPQMTDE 480

QY 481 TRNLQNVDMKIGV 493

DB 481 TRNLQNVDMKIGV 493

RESULT 13

ABU04635 standard; protein; 493 AA.

AC ABU04635;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1301.

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

KM protease; protease inhibitor; transporter; cytoskeletal protein;

KM receptor; transcription factor; cancer; MHC;

KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;

KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

PN WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCOS INC.

PI Chicx RM, Tomlinson AJ, Urban RG;

PT WPI; 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
cytoskeletal proteins, receptors or transcription factors), useful for
treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
leukemia.

PS Example 2; SEQ ID NO 1299; 134pp; English.

CC The invention describes a purified polypeptide, which comprises a
fragment of a kinase, phosphatase, protease, protease inhibitor,
transporter, cytoskeletal protein, receptor or transcription factor. The
polypeptide is useful as an immunogenic composition for eliciting in a
mammal an immunogenic response directed against any of the purified

CC polypeptide. The purified polypeptide, or the antibody that binds to this
polypeptide, is useful for treating cancer. The polypeptide binds also
useful for identifying compounds that binds to a naturally processed

CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing

CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an
expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

DR WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX
PS Example 2; SEQ ID NO 1301; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease or transcription factor. The
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 493 AA:
Query Match 99.3%; Score 2590; DB 6; Length 493;
Best Local Similarity 99.4%; Pred. No. 2.9e-190;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MDKFWHAAWGLCLVPLSLAQIDLNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
DB 1 MDKFWHAAWGLCLVPLSLAQIDLNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
QY 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120
DB 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120
QY 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120
DB 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120
QY 121 ASAPBEDCTSVTLDPNAPDGPITITIVNRDGTTRYVQKGEYRTNPEIDYPSNPTDDVSS 180
DB 121 ASAPBEDCTSVTLDPNAPDGPITITIVNRDGTTRYVQKGEYRTNPEIDYPSNPTDDVSS 180
QY 181 GSSSERSSSTGGYIFTFSTVHPIDPDSFWITSDTRIPATMDSHSTTIOPTANPT 240
DB 181 GSSSERSSSTGGYIFTFSTVHPIDPDSFWITSDTRIPATMDSHSTTIOPTANPT 240
QY 241 GVEDDLPRTPLSMNTQOSNQSFSSTHELEEDKHPTSTLTSNRNDVYGGRRDPNH 300
DB 241 GVEDDLPRTPLSMNTQOSNQSFSSTHELEEDKHPTSTLTSNRNDVYGGRRDPNH 300
QY 301 SEGSTVLEGYTSHYPTKSRPIPTVSAKTSFGVYAVTVGDSNNVRSISGDDTF 360
DB 301 SEGSTVLEGYTSHYPTKSRPIPTVSAKTSFGVYAVTVGDSNNVRSISGDDTF 360
QY 361 HPSGGSHTHGSSSDGSHSGOGGANTTSGPTRTPOIPWLTLLSLALMTLAVCTA 420
DB 361 HPSGGSHTHGSSSDGSHSGOGGANTTSGPTRTPOIPWLTLLSLALMTLAVCTA 420
QY 421 VNSRRRCQKKKLVINGNCAVDRKPSGLNGEASRQEMVHLVNKESSETPDQFTADE 480
DB 421 VNSRRRCQKKKLVINGNCAVDRKPSGLNGEASRQEMVHLVNKESSETPDQFTADE 480
QY 481 TRNLOVDMKIGV 493
DB 481 TRNLOVDMKIGV 493
RESULT 14
ABU04613 standard; protein; 493 AA.

XX
AC ABU04613;
XX
XX 29-JAN-2003 (first entry)
DT
XX
DE Human expressed protein tag (EPT) #1279.
XX
XX Translational profiling; expressed protein tag; EPT; Kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX Homo sapiens.
PN WO200278524-A2.
PD 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US009671.
PP
XX 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.
XX
XX (ZYCO-) ZYCO5 INC.
XX
XX Chicx RM, Tomlinson AJ, Urban RG;
XX
DR WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX
PS Example 2; SEQ ID NO 1279; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease or transcription factor.
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 493 AA:
Query Match 99.3%; Score 2590; DB 6; Length 493;
Best Local Similarity 99.4%; Pred. No. 2.9e-190;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MDKFWHAAWGLCLVPLSLAQIDLNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
DB 1 MDKFWHAAWGLCLVPLSLAQIDLNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL 60
QY 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120
DB 61 PTMAQWEKALSIGETCRGFIIEGHVAVIPRIHNSICANNNGVYILTYNTSQDYTCFN 120

Oy		12	ASAPPEEDCSVYLDLPNADGPIITTTIVNRDSTRVVGGEVRTNPEDIYSNPPTDDVS	180
Dd		121	ASAPPEEDCTSTDLPNADGPITTTIYNRDETRVVGGERINPEDIYSNPPTDDVS	180
Oy		181	GSSSERSSTGGGYIFPFTSTVPHPIDEDSPWITTDSDRIPATNMDSHSHTLQPTANPNT	240
Dd		181	GSSSERSSTSGGIYPFTSFTHVIPIDEDSPWTITDSDRIPATNMDSHSHTLQPTANPNT	240
Oy		241	GLVEDDRLRTPFLSMETQQSNQSQFSFSHEGLEEDDPFTTSLTSSNRNDVTGGRRDPNH	300
Dd		241	GLVEDDLDRCPFLSMETQQSNQSQFSFSHGLEEDDPHTTSLTSSNRNDVTGGRRDPNH	300
Oy		301	SSESSTLLBGTSHYHYHYTESRTPLIPVTSIAKTGSFGTVATVYGDSNVNKSLSGDODTF	360
Dd		301	SEGSTTLBLEYTHYHYHYTERSTPLIPVTSIAKTSFGTVAVTVGDSNVNKSLSGDODTF	360
Oy		361	HPSGSGHTTGSESDSHSGSQEGGANTSGPIRTPOIPENWLIIIASLIALLALILAVCIA	420
Dd		361	HPSGSGHTTGSSDDSHSGSQEGGANTSGPIRKIPOLPENWIIIASLIALLALILAVCIA	420
Oy		421	VNSRRRCQGCKKLVI NSGVAVEDRPXSGELNGEASKSQSEMVLVNKBSSETPDQPMTAD	480
Dd		421	VNSRRRCQGCKKLVI NSGGAVEDRKPESGLNGEASKSQSEMVLVNKBSSETPDQPMTAD	480
Oy		481	TRLQNVDNMKGIV	493
Dd		481	TRLQNVDNMKGIV	493
RESULT 15				
ID	AAR91445			
XX	AAR91445 standard; protein; 493 AA.			
AC	AAR91445:			
XX				
DT	25-MAR-2003	(revised)		
DT	31-OCT-1996	(first entry)		
XX	Human epithelial CD44 antigen.			
XX				
KW	Cell surface antigen; cloning; immunoselection; immunotherapy; therapy;			
KW	diagnosis; vector; epithelial CD44 antigen; carcinoma; cancer;			
KW	metastasis.			
XX				
OS	Homo sapiens.			
XX				
XX				
FT	Key	Location/Qualifiers		
FT	Peptide	1..19		
FT		/label= Sig_peptide		
FT	Domain	20..400		
FT		/label= Extracellular_domain		
FT	Modified-site	65..67		
FT		/label= glycosylation_site		
FT	Modified-site	97..99		
FT		/label= glycosylation_site		
FT	Modified-site	140..142		
FT		/label= glycosylation_site		
FT	Modified-site	150..152		
FT		/label= glycosylation_site		
FT	Modified-site	160..162		
FT		/label= glycosylation_site		
FT	Modified-site	180..182		
FT		/label= glycosylation_site		
FT	Modified-site	190..192		
FT		/label= glycosylation_site		
FT	Modified-site	390..392		
FT		/label= glycosylation_site		
FT	Domain	401..421		
FT		/label= Transmembrane_domain		
FT	Domain	422..493		
FT		/label= Cytoplasmic_domain		
FT	Modified-site	427..429		
FT		/label= glycosylation_site		

[illegible]

Tue Mar 9 08:08:40 2004

pct-us01-51014-1.rag

Page 13

Qy	481	TRNLQNVDMKIGV	493
Db	481	TRNLQNVDMKIGV	493

Search completed: March 8, 2004, 06:10:03
Job time : 62 secs

Tue Mar 9 08:08:41 2004

pct-us01-51014-1.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:10:05 ; Search time 397 Seconds
(without alignments)
262.213 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 1 MDKFWHAAAGCLVPLSLA.....QMTADETRLQNVDMKIV 493

Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: /cn2_6/prodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cn2_6/prodata/2/pubppa/PCT_NEW_PUB.pep:*
3: /cn2_6/prodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cn2_6/prodata/2/pubppa/US06_PUBCOMB.pep:*
5: /cn2_6/prodata/2/pubppa/US07_NEW_PUB.pep:*
6: /cn2_6/prodata/2/pubppa/PCTUS_PUBCOMB.pep:*
7: /cn2_6/prodata/2/pubppa/US08_NEW_PUB.pep:*
8: /cn2_6/prodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
10: /cn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cn2_6/prodata/2/pubppa/US09_NEW_PUB.pep:*
12: /cn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
13: /cn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cn2_6/prodata/2/pubppa/US10_NEW_PUB.pep:*
16: /cn2_6/prodata/2/pubppa/US60_NEW_PUB.pep:*
17: /cn2_6/prodata/2/pubppa/US60_PUBCOMB.pep:*
18: /cn2_6/prodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2489.5	95.5	700	14	US-10-012-969C-2
2	2468.5	94.7	742	10	US-09-983-000A-16
3	2462.5	94.4	742	15	US-10-116-275-206
4	1361	52.2	365	9	US-09-870-759-73
5	1361	52.2	365	10	US-09-751-708A-73
6	1299	48.8	339	14	US-10-291-634-3
7	744	28.5	143	14	US-10-156-932-18
8	734.5	27.2	185	14	US-10-156-932-20
9	718	26.5	187	14	US-10-156-932-22
10	537	20.6	109	9	US-10-156-932-16
11	478	18.3	90	4	US-09-799-118-3
12	473	18.1	90	10	US-09-927-463-8
13	386	14.8	79	14	US-10-156-932-14
14	349	13.4	67	14	US-10-052-641-2
15	199.5	7.6	322	9	US-09-909-320-201

16	199.5	7.6	322	9	US-09-909-088B-201	Sequence 201, App
17	199.5	7.6	322	9	US-09-905-291A-201	Sequence 201, App
18	199.5	7.6	322	9	US-09-902-853-201	Sequence 201, App
19	199.5	7.6	322	9	US-09-907-824-201	Sequence 201, App
20	199.5	7.6	322	9	US-09-907-841-201	Sequence 201, App
21	199.5	7.6	322	10	US-09-904-011-201	Sequence 201, App
22	199.5	7.6	322	10	US-09-906-742-201	Sequence 201, App
23	199.5	7.6	322	10	US-09-906-838-201	Sequence 201, App
24	199.5	7.6	322	10	US-09-907-613-201	Sequence 201, App
25	199.5	7.6	322	10	US-09-904-942-201	Sequence 201, App
26	199.5	7.6	322	10	US-09-904-859-201	Sequence 201, App
27	199.5	7.6	322	10	US-09-909-204-201	Sequence 201, App
28	199.5	7.6	322	10	US-09-904-820-201	Sequence 201, App
29	199.5	7.6	322	10	US-09-904-786-201	Sequence 201, App
30	199.5	7.6	322	10	US-09-906-646-201	Sequence 201, App
31	199.5	7.6	322	10	US-09-906-700-201	Sequence 201, App
32	199.5	7.6	322	10	US-09-903-786-201	Sequence 201, App
33	199.5	7.6	322	10	US-09-902-903-201	Sequence 201, App
34	199.5	7.6	322	10	US-09-903-749A-201	Sequence 201, App
35	199.5	7.6	322	10	US-09-904-956-201	Sequence 201, App
36	199.5	7.6	322	10	US-09-902-736-201	Sequence 201, App
37	199.5	7.6	322	10	US-09-907-794-201	Sequence 201, App
38	199.5	7.6	322	10	US-09-903-943-201	Sequence 201, App
39	199.5	7.6	322	10	US-09-904-462-201	Sequence 201, App
40	199.5	7.6	322	10	US-09-907-925-201	Sequence 201, App
41	199.5	7.6	322	10	US-09-902-692-201	Sequence 201, App
42	199.5	7.6	322	10	US-09-903-520-201	Sequence 201, App
43	199.5	7.6	322	10	US-09-905-056-201	Sequence 201, App
44	199.5	7.6	322	10	US-09-909-064-201	Sequence 201, App
45	199.5	7.6	322	10	US-09-909-064-201	Sequence 201, App

ALIGNMENTS

RESULT 1					
US-10-012-969C-2					
Sequence 2, Application US/10012969C					
Publication No. US20030108984A1					
GENERAL INFORMATION:					
APPLICANT: NAOR, DAVID					
APPLICANT: GOLAN, ITSHAK					
APPLICANT: NEVEZETI, SHOMO					
TITLE OF INVENTION: C944 SPLICE VARIANT ASSOCIATED WITH RHEUMATOID ARTHRITIS					
FILE REFERENCE: C00L21-2					
CURRENT FILING DATE: 2002-09-25					
PRIOR APPLICATION NUMBER: PCT/IL00/00326					
PRIOR FILING DATE: 2000-06-07					
NUMBER OF SEQ ID NOS: 2					
SOFTWARE: PatentIn version 3.1					
SEQ ID NO 2					
LENGTH: 700					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-10-012-969C-2					
Query Match	95.5%	Score 2489.5	DB 14	Length 700	
Best Local Similarity	70.3%	Pred. No. 6.5e-190	Indels 207	Gaps 1	
Matches	492	Conservative	0		
QY	1	MDKFWHAAAGCLVPLSLA	IDNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL	60	
DB	1	MDKFWHAAAGCLVPLSLA	IDNTCRFAGVHVEKNGRYSISRTAADLCAFNSTL	60	
QY	61	PTMAOMERALSIGFTGCGYFIEGHVYPRHPSICAAANTGVYILTSNTSOYDTCCFN	120		
DB	61	PTMAOMERALSIGFTGCGYFIEGHVYPRHPSICAAANTGVYILTSNTSOYDTCCFN	120		
QY	121	ASAPBEDCTSVTDLPNAPDPITITIVNRDGTIRYVKGERTNPEDYPSNPDDVSS	180		
DB	121	ASAPBEDCTSVTDLPNAPDPITITIVNRDGTIRYVKGERTNPEDYPSNPDDVSS	180		


```
QY 181 GSSSERSTSGGYIFTFSTVHPIDPDSWINDSDIRIPAT----- 222
DB 181 GSSSERSTSGGYIFTFSTVHPIDPDSWINDSDIRIPATSTSSNTISAGNEPNEE 240
QY 223 ----- 222
DB 241 DERDRHLSPSSGSDIDDEDPISSSTSTPRAPDHTKONQDWTQWNSHNPVILQTTTR 300
QY 223 ----- 222
DB 301 MTADVDNRGTAYEGNMPNPAAPLIRHEHHEEETPSTSTQATPSTTEETAQOKO 360
QY 223 ----- 222
DB 361 WFGNRMHGVRQTRPREDSSHSTGTGAASAHSHMQGRTPSPEDSSWTDFFNPISHPMG 420
QY 223 -----NDSGSHSTTLOPTANPNTGLVEDLDRGTGPLSNTTQOSNSQSPSTSHGLEE 273
DB 421 RGHQAGRRMDSSHSTTLOPTANPNTGLVEDLDRGTGPLSMTTQOSNSQSPSTSHGLEE 480
QY 274 DKDHPTSTLTSNRNDVYTGRRDPNHSSEGSTTLLGEGTSHYPHTKESRTFIPVTSAKTG 333
DB 481 DKDHPTSTLTSNRNDVYTGRRDPNHSSEGSTTLLGEGTSHYPHTKESRTFIPVTSAKTG 540
QY 334 SPGYTAVTVGDNSNPNVNSLSGDODTTPHSGSHHTTGSGSDGSHSGQEGGANVTSGPI 393
DB 541 SPGYTAVTVGDNSNPNVNSLSGDODTTPHSGSHHTTGSGSDGSHSGQEGGANVTSGPI 600
QY 394 RTPQPEMLILASLIALALILAVCIANSRRCGQKKQVYNSGNAVDRKPSGLNGE 453
DB 601 RTPQPEMLILASLIALALILAVCIANSRRCGQKKQVYNSGNAVDRKPSGLNGE 660
QY 454 ASKSGEMVHLVNKESSETPDQPMTADETRNLOVDMKIGV 493
DB 661 ASKSGEMVHLVNKESSETPDQPMTADETRNLOVDMKIGV 700

RESULT 2
US-09-983-000A-16
; Sequence 16, Application US/09983000A
; Publication No. US20030118585A1
; GENERAL INFORMATION:
; APPLICANT: AGY Therapeutics
; APPLICANT: Mellicker, Sabrine
; APPLICANT: Chiu, Daniel
; TITLE OF INVENTION: USE OF PROTEIN BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALIZ
; TITLE OF INVENTION: OF BRAIN TUMORS
; FILE REFERENCE: 263/180 -- Pegaleman -- AGY
; CURRENT APPLICATION NUMBER: US/09/983,000A
; CURRENT FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 16
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(20)
; OTHER INFORMATION: BY SIMILARITY
; NAME/KEY: Gene
; LOCATION: (1)..(742)
; OTHER INFORMATION: CD-44 Antigen
; NAME/KEY: CHAIN
; LOCATION: (21)..(742)
; OTHER INFORMATION: CD44 ANTIGEN
; NAME/KEY: DOMAIN
; LOCATION: (21)..(649)
; OTHER INFORMATION: EXTRACELLULAR (POTENTIAL)
; NAME/KEY: TRANSMEM
; LOCATION: (650)..(670)
; OTHER INFORMATION: POTENTIAL

NAME/KEY: DOMAIN
LOCATION: (671)..(742)
OTHER INFORMATION: CYTOPLASMIC (POTENTIAL)
NAME/KEY: DOMAIN
LOCATION: (48)..(119)
OTHER INFORMATION: LINK
NAME/KEY: MOD RES
LOCATION: (21)..(21)
OTHER INFORMATION: PYROLIDONE CARBOXYLIC ACID (PROBABLE)
NAME/KEY: DISULFID
LOCATION: (53)..(118)
OTHER INFORMATION: BY SIMILARITY
NAME/KEY: DISULFID
LOCATION: (77)..(97)
OTHER INFORMATION: BY SIMILARITY
NAME/KEY: CARBOHYD
LOCATION: (25)..(25)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (57)..(57)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (100)..(100)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (110)..(110)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (120)..(120)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (350)..(350)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (548)..(548)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (599)..(599)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (636)..(636)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: VARSPPLIC
LOCATION: (23)..(29)
OTHER INFORMATION: DMITCR -> GVGRRKS (IN ISOFORM CD44SP)
NAME/KEY: VARSPPLIC
LOCATION: (30)..(742)
OTHER INFORMATION: MISSING (IN ISOFORM CD44SP)
NAME/KEY: VARSPPLIC
LOCATION: (192)..(192)
OTHER INFORMATION: G -> A (IN ISOFORM WITH ALTERNATIVE SPLICE DONOR/ACCEPTOR ON
NAME/KEY: VARSPPLIC
LOCATION: (193)..(223)
OTHER INFORMATION: MISSING (IN ISOFORM WITH ALTERNATIVE SPLICE DONOR/ACCEPTOR O
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> S (IN ISOFORM WITHOUT EXON 6)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(266)
OTHER INFORMATION: MISSING (IN ISOFORM WITHOUT EXON 6)
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> N (IN ISOFORMS WITHOUT EXONS 6 TO 11)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(472)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXONS 6 TO 11)
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> R (IN ISOFORMS WITHOUT EXONS 6 TO 14)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(604)
```

```

/ OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXONS 6 TO 14)
/ NAME/KEY: VARSPLIC
/ LOCATION: (266) ..(273)
/ OTHER INFORMATION: MISSING (IN ISOFORM WITH ALTERNATIVE SPLIC DONOR/ACCEPTOR ON EXC
/ OTHER INFORMATION: N 7)
/ NAME/KEY: VARSPLIC
/ LOCATION: (385) ..(385)
/ OTHER INFORMATION: I -> T (IN ISOFORMS WITHOUT EXON 10)
/ NAME/KEY: VARSPLIC
/ LOCATION: (386) ..(428)
/ OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 10)
/ NAME/KEY: VARSPLIC
/ LOCATION: (506) ..(506)
/ OTHER INFORMATION: Q -> R (IN ISOFORMS WITHOUT EXON 13)
/ NAME/KEY: VARSPLIC
/ LOCATION: (507) ..(535)
/ OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 13)
/ NAME/KEY: VARSPLIC
/ LOCATION: (536) ..(536)
/ OTHER INFORMATION: N -> R (IN ISOFORMS WITHOUT EXON 14)
/ NAME/KEY: VARSPLIC
/ LOCATION: (537) ..(604)
/ OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 14)
/ NAME/KEY: VARSPLIC
/ LOCATION: (675) ..(675)
/ OTHER INFORMATION: R -> S (IN ISOFORMS WITHOUT EXON 19)
/ NAME/KEY: VARSPLIC
/ LOCATION: (676) ..(742)
/ OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 19)
/ NAME/KEY: VARIANT
/ LOCATION: (46) ..(46)
/ OTHER INFORMATION: R -> P (IN IN(A) ANTIGEN)
/ US-09-983-000A-16

```

Query Match 94.7%; Score 2468.5; DB 10; Length 742;
 Best Local Similarity 66.3%; Pred. No. 3.3e-188;
 Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

```

Qy 1 MDKFWHAAWGLCLVPLSLAQIDLNITCRFAGVFNKGRYSISRTAADLCKAFNSTL 60
Db 1 MDKFWHAAWGLCLVPLSLAQIDLNITCRFAGVFNKGRYSISRTAADLCKAFNSTL 60
Qy 61 PTMAQMEKALISIGETCRYGIFBHVIPRIHNSICAANNVGYLLTNTSOYDTYCFN 120
Db 61 PTMAQMEKALISIGETCRYGIFBHVIPRIHNSICAANNVGYLLTNTSOYDTYCFN 120
Qy 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
Qy 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
Db 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
Qy 223 ----- 222
Db 223 ----- 222
Qy 241 TWDWFSWFLPSESXKHLHTTQMAGTSNTISAGWEPNENEDERDRHLSFGSGIDDD 300
Db 241 TWDWFSWFLPSESXKHLHTTQMAGTSNTISAGWEPNENEDERDRHLSFGSGIDDD 300
Qy 223 ----- 222
Db 223 ----- 222
Qy 301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTDVRNGTAYEGNMN 360
Db 301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTDVRNGTAYEGNMN 360
Qy 223 ----- 222
Db 223 ----- 222
Qy 421 HSTTGTAASAHSHPMQGRTPSPEDSSWTDFNPISHPMGRGHQAGRRMMDSSHTT 480
Db 421 HSTTGTAASAHSHPMQGRTPSPEDSSWTDFNPISHPMGRGHQAGRRMMDSSHTT 480
Qy 232 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 291
Db 232 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 291
Qy 481 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 540
Db 481 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 540

```

```

Qy 292 TGRGRDNHSEGSTLLLEGYTSHPYTKESRTFIPYTSATGSEFGVAVTVGDSNSVNR 351
Db 541 TGRGRDNHSEGSTLLLEGYTSHPYTKESRTFIPYTSATGSEFGVAVTVGDSNSVNR 600
Qy 352 SLSDQDTEFHPGSHHTHSGSDGSHSGOEGANTTSGPIRTPQIPENLILASLAL 411
Db 601 SLSDQDTEFHPGSHHTHSGSDGSHSGOEGANTTSGPIRTPQIPENLILASLAL 660
Qy 412 ALIIVAVNRRRCQKKLYINSNGAVEDRKPSGLNGEASKSQEMVHLVVKESSET 471
Db 661 ALIIVAVNRRRCQKKLYINSNGAVEDRKPSGLNGEASKSQEMVHLVVKESSET 720
Qy 472 PDQFMTADETRNLQNDMKIGV 493
Db 721 PDQFMTADETRNLQNDMKIGV 742

```

RESULT 3

```

US-10-116-275-206
/ Sequence 206, Application US/10116275
/ Publication No. US2003021476A1
/ GENERAL INFORMATION:
/ APPLICANT: Elian Pharmaceutical Technology
/ APPLICANT: O'Mahony, Daniel J.
/ APPLICANT: Brayden, David
/ APPLICANT: Byrne, Daragh
/ APPLICANT: Lambkin, Imelda
/ APPLICANT: Higgins, Lisa
/ TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods
/ FILE REFERENCE: E1067/20087
/ CURRENT APPLICATION NUMBER: US/10/116,275
/ CURRENT FILING DATE: 2002-10-04
/ NUMBER OF SEQ ID NOS: 349
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 206
/ LENGTH: 742
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-116-275-206

```

Query Match 94.4%; Score 2462.5; DB 15; Length 742;
 Best Local Similarity 66.2%; Pred. No. 1e-187;
 Matches 491; Conservative 1; Mismatches 1; Indels 249; Gaps 1;

```

Qy 1 MDKFWHAAWGLCLVPLSLAQIDLNITCRFAGVFNKGRYSISRTAADLCKAFNSTL 60
Db 1 MDKFWHAAWGLCLVPLSLAQIDLNITCRFAGVFNKGRYSISRTAADLCKAFNSTL 60
Qy 61 PTMAQMEKALISIGETCRYGIFBHVIPRIHNSICAANNVGYLLTNTSOYDTYCFN 120
Db 61 PTMAQMEKALISIGETCRYGIFBHVIPRIHNSICAANNVGYLLTNTSOYDTYCFN 120
Qy 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDDVSS 180
Qy 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
Db 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
Qy 223 ----- 222
Db 223 ----- 222
Qy 241 TWDWFSWFLPSESXKHLHTTQMAGTSNTISAGWEPNENEDERDRHLSFGSGIDDD 300
Db 241 TWDWFSWFLPSESXKHLHTTQMAGTSNTISAGWEPNENEDERDRHLSFGSGIDDD 300
Qy 223 ----- 222
Db 223 ----- 222
Qy 301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTDVRNGTAYEGNMN 360
Db 301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTDVRNGTAYEGNMN 360
Qy 223 ----- 222
Db 223 ----- 222
Qy 421 HSTTGTAASAHSHPMQGRTPSPEDSSWTDFNPISHPMGRGHQAGRRMMDSSHTT 480
Db 421 HSTTGTAASAHSHPMQGRTPSPEDSSWTDFNPISHPMGRGHQAGRRMMDSSHTT 480
Qy 232 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 291
Db 232 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 291
Qy 481 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 540
Db 481 LQTPANNTGLVNDLDRGTPLSMTOOSNOSFSTSEGLLEDKHPTTSTLTSRNRY 540

```

QY 223 -----NMDSHSTT 231
DB 421 HSTGTAAASHTSHPMQGRTPSPEDSWTFNFI SHPMGRHQAGRMMDSSHIT 480
QY 232 LQIPANPNTGLVBDLTGTPLSMTTQOSNSQSFSTHGELEEDKHPTTSTLSSNRNV 291
DB 481 LQIPANPNTGLVBDLTGTPLSMTTQOSNSQSFSTHGELEEDKHPTTSTLSSNRNV 540
QY 292 TGRARDPNHSGSTTLLEGYTSYHPTKESRTFIPYTSAKTSGFVTAATVGDNSNVNR 351
DB 541 TGRARDPNHSGSTTLLEGYTSYHPTKESRTFIPYTSAKTSGFVTAATVGDNSNVNR 600
QY 352 SLGSDOTFHPGSGSHTHGSESDHSGSQEGANTTSGPITPQIPFWLLILASLAL 411
DB 601 SLGSDOTFHPGSGSHTHGSESDHSGSQEGANTTSGPITPQIPFWLLILASLAL 660
QY 412 ALLAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESST 471
DB 661 ALLAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESST 720
QY 472 PDQPMTADETRNLQNDMKIGV 493
DB 721 PDQPMTADETRNLQNDMKIGV 742

RESULT 4

US-09-870-759-73
; Sequence 73, Application US/09870759
; Patent No. US2002017551A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 365
; TYPE: PRF
; ORGANISM: Mus musculus
US-09-870-759-73

Query Match 52.2%; Score 1361; DB 9; Length 365;
Best Local Similarity 57.7%; Pred. No. 2.8e-100;
Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

QY 1 MDKFWMTAAWGLCLVPLSLA-----QIDLNTCPAGVFHYEKNRGYSISRTAADLCKAF 56
DB 1 MDKFWMTAAWGLCLQSLAHPHQOIDLNTCYRAGVFHYEKNRGYSISRTAADLCKAF 60
QY 57 NSTLPTMAOMEKALSIGETCRVGFIEGHVVIPIRHNSICAANTGVYL--TSNTSOYD 115
DB 61 NSTLPTMDQKMLASKGETCRVGFIEGHVVIPIRHNAICANHGVILVTSNTSHYD 120
QY 116 TYCFNAPAPREDCTSVTLDPNAPDGPITITIVNRDGRYVQKGEYRTNEDIYPSNPTD 175
DB 121 TYCFNAPAPREDCTSVTLDPNAPDGPITITIVNRDGRYVQKGEYRTNEDIYPSNPTD 180
QY 176 DVVSGSSERSSTSGGYIFTFSTVHPIDEDSPWITDSTRIPATNMSSHTLOPT 235
DB 181 DVVSGSTIEK-STPESTILHTY-----LPTE-----QPT 209
QY 236 ANPNTGLVEDLDRGTPLSMTTQOSNSQSFSTHGELEEDKHPTTSTLSSNRNDVTGGR 295
DB 210 G-----DQDSSFIRSTLATRDSDS----- 229
QY 296 RDPNHSSESTTLLEGYTSYHPTKESRTFIPYTSAKTSGFVTAATVGDNSNVNLSLG 355
DB 230 -----SKDSR----- 234

QY 356 DQDTHPSGGSHT--THGSESDHSGSQEGANTTSGPITPQIPFWLLILASLALALI 414
DB 225 -----GSSRTVTHGSELGHSSANQDGVTTTSGPMRPPQIPFWLLILASLALALI 286
QY 415 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESSTPDD 474
DB 287 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESSTPDD 346
QY 475 FMTADETRNLQNDMKIGV 493
DB 347 FMTADETRNLQNDMKIGV 365

RESULT 5

US-09-751-708A-73
; Sequence 73, Application US/09751708A
; Publication No. US20030157113A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 751708
; CURRENT APPLICATION NUMBER: US/09/751,708A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/173,371
; PRIOR FILING DATE: 1999-12-28
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 365
; TYPE: PRF
; ORGANISM: Mus musculus
US-09-751-708A-73

Query Match 52.2%; Score 1361; DB 10; Length 365;
Best Local Similarity 57.7%; Pred. No. 2.8e-100;
Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

QY 1 MDKFWMTAAWGLCLVPLSLA-----QIDLNTCPAGVFHYEKNRGYSISRTAADLCKAF 56
DB 1 MDKFWMTAAWGLCLQSLAHPHQOIDLNTCYRAGVFHYEKNRGYSISRTAADLCKAF 60
QY 57 NSTLPTMAOMEKALSIGETCRVGFIEGHVVIPIRHNSICAANTGVYL--TSNTSOYD 115
DB 61 NSTLPTMDQKMLASKGETCRVGFIEGHVVIPIRHNAICANHGVILVTSNTSHYD 120
QY 116 TYCFNAPAPREDCTSVTLDPNAPDGPITITIVNRDGRYVQKGEYRTNEDIYPSNPTD 175
DB 121 TYCFNAPAPREDCTSVTLDPNAPDGPITITIVNRDGRYVQKGEYRTNEDIYPSNPTD 180
QY 176 DVVSGSSERSSTSGGYIFTFSTVHPIDEDSPWITDSTRIPATNMSSHTLOPT 235
DB 181 DVVSGSTIEK-STPESTILHTY-----LPTE-----QPT 209
QY 236 ANPNTGLVEDLDRGTPLSMTTQOSNSQSFSTHGELEEDKHPTTSTLSSNRNDVTGGR 295
DB 210 G-----DQDSSFIRSTLATRDSDS----- 229
QY 296 RDPNHSSESTTLLEGYTSYHPTKESRTFIPYTSAKTSGFVTAATVGDNSNVNLSLG 355
DB 230 -----SKDSR----- 234
QY 356 DQDTHPSGGSHT--THGSESDHSGSQEGANTTSGPITPQIPFWLLILASLALALI 414
DB 225 -----GSSRTVTHGSELGHSSANQDGVTTTSGPMRPPQIPFWLLILASLALALI 286
QY 415 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESSTPDD 474
DB 287 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNKESSTPDD 346
QY 475 FMTADETRNLQNDMKIGV 493
DB 347 FMTADETRNLQNDMKIGV 365

RESULT 6
US-10-291-634-3
; Sequence 3, Application US/10291634
; Publication No. US20030105058A1
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L
; APPLICANT: Ni, Jian
; APPLICANT: Dillion, Patrick J
; TITLE OF INVENTION: CD-44 Like Protein
; FILE REFERENCE: 1488, 0490003
; CURRENT APPLICATION NUMBER: US/10/291,634
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/288,230
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/892,880
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/021,762
; PRIOR FILING DATE: 1996-07-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-291-634-3
Query Match 49.8%; Score 1299; DB 14; Length 339;
Best Local Similarity 56.4%; Pred. No. 2.3e-95;
Matches 279; Conservative 18; Mismatches 38; Indels 160; Gaps 8;
QY 1 MDKFWHAANG-LCLVLSLA--QIDNITCRPAQVHHKNGKRGISRTAADLCKAN 57
DB 1 MDKFWHAANG-LCLVLSLA--QIDNITCRPAQVHHKNGKRGISRTAADLCKAN 60
QY 58 STPTMAQMEKALSIGETCRGFGHGVVPIPHPSICANNQGYI-LTSNTSQYOT 116
DB 61 TITPTMAQMEKALRKGETCRGFGHGVVPIPHPSICANNQGYIILASNTSHYT 120
QY 117 YCFNASAPBEDCTSVTDLPNADGPITITIVNRDGRYVQKGYRTNPEDIYPSNPTD 176
DB 121 YCFNASAPBEDCTSVTDLPNADGPITITIVNRDGRYVQKGYRTNPEDIYPSNPTD 180
QY 177 DVSSGSSERSSTGCIYFTFTSVHPIPEDCSPMITDSDRIIPANMSSHHTLOPTA 236
DB 181 DVSSGSSERSSTGCIYFTFTSVHPIPEDCSPMITDSDRIIPANMSSHHTLOPTA 208
QY 237 NPMTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKHPITSTLSSNRNDVTGGR 296
DB 209 -----GDR 211
QY 297 DPNHSGSTTLLEGYTHYHPTKESRTFIPVTSKATGSPGVAVVYVGDNSNVNRLSGD 356
DB 212 DDAFFIGSTL----- 221
QY 357 QDTFHPGSGHTTHGSDHSHSQSGGANTTSGPIRTQIPEMLIILASLALALILA 416
DB 222 -----ATGHSSGNQDSGVTTTSGPARRRQIPEMLIILASLALALILA 264
QY 417 VCLAVNSRRRCQKKKLVNSGCAVEDRPSGLNGEASQSCMVLNKKESERTDQDM 476
DB 265 VCLAVNSRRRCQKKKLVNSGCAVEDRPSGLNGEASQSCMVLNKKESERTDQDM 324
QY 477 TADETRNLQNVDMKI 491
DB 325 TADETRNLQNVDMKI 339

RESULT 7
US-10-156-932-18
; Sequence 18, Application US/10156932
; Publication No. US20030069181A1

GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-18
Query Match 28.5%; Score 744; DB 14; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.3e-51;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 217 DRIPATNDSHSTTLOPTANPNTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKD 276
DB 1 DRIPATNDSHSTTLOPTANPNTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKD 60
QY 277 HPTTSTLSSNRNDVYTGGRDPNHSSEGSTTLLEGYTSHPHTKESRTFIPVTSKATGSPG 336
DB 61 HPTTSTLSSNRNDVYTGGRDPNHSSEGSTTLLEGYTSHPHTKESRTFIPVTSKATGSPG 120
QY 337 VTAVTVGDSNNSVNRSLSGDDOT 359
DB 121 VTAVTVGDSNNSVNRSLSGDDOT 143
RESULT 8
US-10-156-932-20
; Sequence 20, Application US/10156932
; Publication No. US20030069181A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-20
Query Match 28.2%; Score 734.5; DB 14; Length 185;
Best Local Similarity 76.0%; Pred. No. 1.1e-50;
Matches 149; Conservative 9; Mismatches 15; Indels 23; Gaps 2;
QY 176 DVSSGSSERSSTGCIYFTFTSVHPIPEDS-----PWIDSDTRIPATN 223
DB 1 DRIPATNDSHSTTLOPTANPNTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKD 49
QY 224 MDSSHSTTLOPTANPNTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKDHPITSTL 283
DB 50 MDSSHSTTLOPTANPNTGLVEDLDRGTPLSMITQOQNSQSFSTSHGLEBDKDHPITSTL 109
QY 284 TSSNRNDVYTGGRDPNHSSEGSTTLLEGYTSHPHTKESRTFIPVTSKATGSPGVATVNG 343
DB 110 TSSNRNDVYTGGRDPNHSSEGSTTLLEGYTSHPHTKESRTFIPVTSKATGSPGVATVNG 169
QY 344 DSNNSVNRSLSGDDOT 359

Db 170 DSNNSNNRSLSGDQT 185

RESULT 9

US-10-156-932-22
; Sequence 22, Application US/10156932
; Publication No. US20030069181A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; TITLE OF INVENTION: Basis for Multiple Therapeutic Modalities
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-22

Query Match 27.5%; Score 718; DB 14; Length 187;
Best Local Similarity 82.5%; Pred. No. 2,2e-49;
Matches 146; Conservative 3; Mismatches 10; Indels 18; Gaps 3;

QY 200 TVHPI-----PDDSPWINDSDIRP-----ATNMDSHSTTLQPTANPTQL 242
DB 12 TSHWQGRTPSPEDSW-TDFEPLSHPMGRGHQAGRMDSSHTILOPTANPTGL 70
QY 243 VEDLDRTGPLSMTQOSNSQSFSTSHGLEEDKDHPPTSTLTSSNRNDVTGGRDPNHS 302
DB 71 VEDLDRTGPLSMTQOSNSQSFSTSHGLEEDKDHPPTSTLTSSNRNDVTGGRDPNHS 130
QY 303 GSTTLLEGYSHYHTESRTFIPVTSKTSFGVAVTVGDSNNSNNRSLSGDQT 359
DB 131 GSTTLLEGYSHYHTESRTFIPVTSKTSFGVAVTVGDSNNSNNRSLSGDQT 187

RESULT 10

US-10-156-932-16
; Sequence 16, Application US/10156932
; Publication No. US20030069181A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; TITLE OF INVENTION: Basis for Multiple Therapeutic Modalities
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-16

Query Match 20.6%; Score 537; DB 14; Length 109;
Best Local Similarity 99.0%; Pred. No. 2,9e-35;
Matches 103; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 256 TQOSNSQSFSTSHGLEEDKDHPPTSTLTSSNRNDVTGGRDPNHSGSTTLLEGYSHY 315
DB 6 TKOSNSQSFSTSHGLEEDKDHPPTSTLTSSNRNDVTGGRDPNHSGSTTLLEGYSHY 65
QY 316 PHTKESRTFIPVTSKTSFGVAVTVGDSNNSNNRSLSGDQT 359
DB 66 PHTKESRTFIPVTSKTSFGVAVTVGDSNNSNNRSLSGDQT 109

RESULT 11

US-09-799-118-3
; Sequence 3, Application US/09799118
; Patent No. US20020090708A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Ho

Wislenski, Hans Georg
Vilcek, Jan

TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
Therefor and Uses Thereof

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESS: Broadway and Neilmark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington

STATE: D.C.
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/799,118

FILING DATE: 06-Mar-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/242,097

FILING DATE: 13-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: LEE26/VILCEK=1B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-628-5197

TELEFAX: 212-737-3528

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 90 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-799-118-3

Query Match 18.3%; Score 478; DB 9; Length 90;
Best Local Similarity 98.9%; Pred. No. 1,1e-30;
Matches 89; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 32 GVEHYENKGRYSISTEADLCXAFNSTLPTMAOMEXALSIGFECRGYFIEGHVPIPI 91
DB 1 GFVHEKNGRYSISTEADLCXAFNSTLPTMAOMEXALSIGFECRGYFIEGHVPIPI 60
QY 92 HENSICAAANTGVYILTSNTSQYDTYCFA 121
DB 61 HENSICAAANTGVYILTSNTSQYDTYCFA 90

RESULT 12

US-09-927-463-8
; Sequence 8, Application US/09927463
; Publication No. US20030032621A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Means for regulating hematopoietic differentiation
; FILE REFERENCE: 1113
; CURRENT APPLICATION NUMBER: US/09/927,463
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1

```
; SEQ ID NO 8
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-927-463-8

Query Match
Best Local Similarity 18.1%; Score 473; DB 10; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.8e-30;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LCVPLSLAQIDNITCRFAGVFRVKNGRYSISRTEADLCKAFSTPTMAQWEKALS 71
DB 1 LCVPLSLAQIDNITCRFAGVFRVKNGRYSISRTEADLCKAFSTPTMAQWEKALS 60

QY 72 IGFETCRVGFIEGHVVIPIRIHPSICANN 101
DB 61 IGFETCRVGFIEGHVVIPIRIHPSICANN 90

RESULT 13
US-10-156-932-14
; Sequence 14, Application US/10156932
; Publication No. US2003006916A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-14

Query Match
Best Local Similarity 14.8%; Score 386; DB 14; Length 79;
Best Local Similarity 94.8%; Pred. No. 2e-23;
Matches 73; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 283 LITSNRDVYGRDRPHSGSTLLRGYSHYPTKESRPTPTSAKTSFGYTAATV 342
DB 3 LITSNRDVYGRDRPHSGSTLLRGYSHYPTKESRPTPTSAKTSFGYTAATV 62

QY 343 GDSNSVNRSLSGDDPT 359
DB 63 GDSNSVNRSLSGDDPT 79

RESULT 14
US-10-052-641-2
; Sequence 2, Application US/10052641
; Publication No. US20030032073A1
; GENERAL INFORMATION:
; APPLICANT: HEIDER, et al.
; TITLE OF INVENTION: Method for Diagnosis and Therapy of Hodgkin Lymphomas
; FILE REFERENCE: 0652.1910000/RSP/KV
; CURRENT APPLICATION NUMBER: US/10/052,641
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: 09/331,254
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: PCT/EP97/07081
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: DE 196 53 607.3
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 67
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-641-2

Query Match
Best Local Similarity 13.4%; Score 349; DB 14; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.4e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 RNDVYGRDRPHSGSTLLRGYSHYPTKESRPTPTSAKTSFGYTAATVGDNS 347
DB 1 RNDVYGRDRPHSGSTLLRGYSHYPTKESRPTPTSAKTSFGYTAATVGDNS 60

QY 348 NVNRSLS 354
DB 61 NVNRSLS 67

RESULT 15
US-09-909-320-201
; Sequence 201, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Baton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie F.
; APPLICANT: Pan, James
; APPLICANT: Paonli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Williams, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
```

Tue Mar 9 08:08:41 2004

pct-us01-51014-1.rabb

Page 8

? PRIOR FILING DATE: 1999-11-30
 ? PRIOR APPLICATION NUMBER: PCT/US99/28564
 ? PRIOR FILING DATE: 1999-12-02
 ? PRIOR FILING DATE: 1999-12-02
 ? PRIOR APPLICATION NUMBER: PCT/US99/28555
 ? PRIOR FILING DATE: 1999-12-02
 ? PRIOR APPLICATION NUMBER: PCT/US99/30095
 ? PRIOR FILING DATE: 1999-12-16
 ? PRIOR APPLICATION NUMBER: PCT/US99/30911
 ? PRIOR FILING DATE: 1999-12-20
 ? PRIOR APPLICATION NUMBER: PCT/US99/30999
 ? PRIOR FILING DATE: 1999-12-20
 ? PRIOR APPLICATION NUMBER: PCT/US00/00219
 ? PRIOR FILING DATE: 2000-01-05
 ? NUMBER OF SEQ ID NOS: 423
 ? SEQ ID NO 201
 ? LENGTH: 322
 ? TYPE: PRT
 ? ORGANISM: Artificial sequence
 ? FEATURE:
 ? OTHER INFORMATION: Synthetic protein
 US-09-909-320-201

Query Match	7.6%;	Score 199.5;	DB 9;	Length 322;
Best Local Similarity	22.2%;	Pred. No. 1.1e-07;		
Matches	83;	Conservative	48;	Mismatches 138;
			Indels 105;	Gaps 12;

Gy		1	LSTAOIDLNTGCFACVFEVEKNGRYSISREADLCGAFNSTLPTVAQOEKLISGFET	76
Dd		25	LRAEELISOVSCHIMGITVLVSKKANQOLNPEBAEKACHLLGLSLAGDQVETLYLKASFET	84
Gy		77	CRYGFL-EGHVLPRIHPNSICANNNGVYLITSNTS-QVDYTCFENASAPDEEDCTSVD	134
Dd		85	CSYGWODGGVVIRISIPMKCKRGVGVLIMKVPSHOFAPAACYNSSDPTWSCI----	140
Gy		135	LPMAFDEPITIIVNRDGTRVQKEERYKTNEPDILPSMFPDDVDVSSSSSEBSSTSOGVI	194
Dd		141	-----PELIITTKXPIFNTOJTAQTTE-----FI	163
Gy		195	FYTFSVTVPIDEDSPMITDSTDRLPATNMDSHSHTLOPANFNTGLVEDLDRTGPLSM	254
Dd		164	-----VSDST-----YSVASPYSTIAPFTTTPPAASISIRKKKLIC	201
Gy		255	TTOO-SNSQSFSHSGLEBDXDHPTTSTLTSSNRNDVYGRADPNHSEGSITLL-----	308
Dd		202	VTEFMETSMTSETEPVENK-----AAFNEMAAAGGVF-----TALLVIALLL	246
Gy		309	-----EGTSHYPHTKESRTIIPYLSAKTOSFGVTLAVGDGNSNVMSLSIG	355
Dd		247	FFGMAALGFCYVKRVYKAAPPFTNNQOKEMLETXY-----VKEEKANDSNFN-EESKTI	300
Gy		356	DODTFHPSGGSHTT	369
Dd		301	DKNPBESKSPTSITT	314

Search completed: March 8, 2004, 06:22:12
Job time : 398 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 23 Seconds
(without alignments)

1106.591 Million cell updates/sec

Title: PCT-US01-51014-1
Perfect score: 2608
Sequence: 1 MDKFWHAAWGLVPLSLA.....QFMADETRNLDVDMKIGV 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1827	70.1	361	1	US-07-946-497-6 Sequence 6, Appli
2	1827	70.1	361	1	US-08-483-322-6 Sequence 6, Appli
3	1827	70.1	361	2	US-08-478-882-6 Sequence 6, Appli
4	1698.5	65.1	362	6	5504194-2 Patent No. 5504194
5	1425	54.6	503	1	US-07-946-497-2 Sequence 2, Appli
6	1425	54.6	503	1	US-08-483-322-2 Sequence 2, Appli
7	1425	54.6	503	2	US-08-478-882-2 Sequence 2, Appli
8	1357	52.0	363	1	US-07-946-497-7 Sequence 7, Appli
9	1357	52.0	363	1	US-08-483-322-7 Sequence 7, Appli
10	1357	52.0	363	2	US-08-478-882-7 Sequence 7, Appli
11	1299	49.8	339	2	US-08-892-880-3 Sequence 3, Appli
12	711	27.3	354	1	US-07-946-497-4 Sequence 4, Appli
13	711	27.3	354	1	US-08-483-322-4 Sequence 4, Appli
14	711	27.3	354	2	US-08-478-882-4 Sequence 4, Appli
15	699	26.8	338	2	US-08-359-850-2 Sequence 2, Appli
16	478	18.3	90	2	US-08-242-097-3 Sequence 3, Appli
17	478	18.3	90	2	US-09-206-695-3 Sequence 3, Appli
18	478	18.3	90	4	US-09-799-118-3 Sequence 3, Appli
19	373	14.3	355	1	US-07-946-497-5 Sequence 5, Appli
20	373	14.3	355	1	US-08-483-322-5 Sequence 5, Appli
21	373	14.3	355	2	US-08-478-882-5 Sequence 5, Appli
22	362	13.9	334	2	US-08-359-850-4 Sequence 4, Appli
23	349	13.4	67	4	US-09-331-254-2 Sequence 2, Appli
24	226.5	8.7	41	2	US-08-143-311B-25 Sequence 25, Appli
25	209	8.0	42	2	US-08-143-311B-4 Sequence 4, Appli
26	209	8.0	42	4	US-08-753-851-4 Sequence 4, Appli
27	200	7.7	318	4	US-09-724-864-60 Sequence 60, Appli

28	199.5	7.6	322	4	US-09-232-160-21	Sequence 21, Appli
29	199.5	7.6	322	4	US-09-907-794A-201	Sequence 201, App
30	199.5	7.6	322	4	US-09-805-125R-201	Sequence 201, App
31	199.5	7.6	322	4	US-09-802-775A-201	Sequence 201, App
32	193.5	7.4	322	2	US-08-892-880-2	Sequence 9, Appli
33	187	7.2	78	3	US-08-564-225-9	Sequence 8, Appli
34	176	6.7	36	3	US-08-564-225-8	Sequence 8, Appli
35	174	6.7	36	2	US-08-143-311B-8	Sequence 8, Appli
36	170	6.5	32	2	US-08-143-311B-3	Sequence 3, Appli
37	170	6.5	32	2	US-08-753-851-3	Sequence 3, Appli
38	167.5	6.4	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
39	164	6.3	31	2	US-08-143-311B-10	Sequence 10, Appli
40	164	6.3	31	4	US-08-753-851-12	Sequence 12, Appli
41	162.5	6.2	277	1	US-08-024-868-2	Sequence 2, Appli
42	162.5	6.2	277	2	US-08-242-097-2	Sequence 2, Appli
43	162.5	6.2	277	3	US-09-206-695-2	Sequence 2, Appli
44	162.5	6.2	277	4	US-09-000-179-1	Sequence 1, Appli
45	162.5	6.2	277	4	US-09-799-118-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-07-946-497-6
Sequence 6, Application US/07946497
Patent No. 5506119
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUNTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: hCD44
US-07-946-497-6
Query Match 70.1%; Score 1827; DB 1; Length 361;
Best Local Similarity 72.6%; Pred. No. 3,6e-156;
Matches 358; Conservative 0; Mismatches 3; Indels 132; Gaps 1;

QY 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVFEKNGRYSISRTAADLCAFNSTL 60
 DB 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVFEKNGRYSISRTAADLCAFNSTL 60
 QY 61 PTMAQMEKALSIGETCRYGFIIEGHVVIPIRHPNSICAAANTGVYLLTSNTSQYDTYCFN 120
 DB 61 PTMAQMEKALSIGETCRYGFIIEGHVVIPIRHPNSICAAANTGVYLLTSNTSQYDTYCFN 120
 QY 121 ASAPPEEDCTSVTDLPAFADGPITITTVNRDGRTRYOKGEYRNPEDIYPSNPPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLPAFADGPITITTVNRDGRTRYOKGEYRNPEDIYPSNPPTDDVSS 180
 QY 181 GSSSERSTSGGYFYFTSTVHPIDPDSPMITDSTRIPATNMDSHSTTLOPTANPT 240
 DB 181 GSSSERSTSGGYFYFTSTVHPIDPDSPMITDSTRIPATNMDSHSTTLOPTANPT 223
 QY 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKHPPTSTLTSSNRNDVTGGRDPNH 300
 DB 224 ----- 223
 QY 301 SEGSTLLEGYTSHPHTKESRTFIPVTSATGSPGVTAVTGDSNVSNSLSGDDPTF 360
 DB 224 -----DDTF 228
 QY 361 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALALILAVCIA 420
 DB 229 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALALILAVCIA 288
 QY 421 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDDQFMTADE 480
 DB 289 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDDQFMTADE 348
 QY 481 TRNLQNVDMKIGV 493
 DB 349 TRNLQNVDMKIGV 361

RESULT 2

US-08-483-322-6

Sequence 6, Application US/08483322

Patent No. 5760178

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter

APPLICANT: PONTA, Helmut

APPLICANT: GUENTHER, Ursula

APPLICANT: WATZKU, Siegfried

APPLICANT: WENZL, Achim

TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,

TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,322

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 07/946,497

FILING DATE: 09-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 16915/145

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 amino acids

TYPE: amino acid

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: hCD44

US-08-483-322-6

Query Match 70.1%; Score 1827; DB 1; Length 361;

Best Local Similarity 72.6%; Pred. No. 3.6e-156;

Matches 358; Conservative 0; Mismatches 3; Indels 132; Gaps 1;

QY 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVFEKNGRYSISRTAADLCAFNSTL 60
 DB 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVFEKNGRYSISRTAADLCAFNSTL 60
 QY 61 PTMAQMEKALSIGETCRYGFIIEGHVVIPIRHPNSICAAANTGVYLLTSNTSQYDTYCFN 120
 DB 61 PTMAQMEKALSIGETCRYGFIIEGHVVIPIRHPNSICAAANTGVYLLTSNTSQYDTYCFN 120
 QY 121 ASAPPEEDCTSVTDLPAFADGPITITTVNRDGRTRYOKGEYRNPEDIYPSNPPTDDVSS 180
 DB 121 ASAPPEEDCTSVTDLPAFADGPITITTVNRDGRTRYOKGEYRNPEDIYPSNPPTDDVSS 180
 QY 181 GSSSERSTSGGYFYFTSTVHPIDPDSPMITDSTRIPATNMDSHSTTLOPTANPT 240
 DB 181 GSSSERSTSGGYFYFTSTVHPIDPDSPMITDSTRIPATNMDSHSTTLOPTANPT 223
 QY 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKHPPTSTLTSSNRNDVTGGRDPNH 300
 DB 224 ----- 223
 QY 301 SEGSTLLEGYTSHPHTKESRTFIPVTSATGSPGVTAVTGDSNVSNSLSGDDPTF 360
 DB 224 -----DDTF 228
 QY 361 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALALILAVCIA 420
 DB 229 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALALILAVCIA 288
 QY 421 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDDQFMTADE 480
 DB 289 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDDQFMTADE 348
 QY 481 TRNLQNVDMKIGV 493
 DB 349 TRNLQNVDMKIGV 361

RESULT 3

US-08-478-882-6

Sequence 6, Application US/08478882

Patent No. 5885575

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter

APPLICANT: PONTA, Helmut

APPLICANT: GUENTHER, Ursula

APPLICANT: WATZKU, Siegfried

APPLICANT: WENZL, Achim

TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,

TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

```

; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/06/478,882
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/946,497
; FILING DATE: 1992.11.09
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16915/145
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ. ID NO. 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: hcd44
; US-06-478-882-6

```

```

Query Match 70.1%; Score 1827; DB 2; Length 361;
Best Local Similarity 72.6%; Pred. No. 3,6e-156;
Matches 358; Conservative 0; Mismatches 3; Indels 122; Gaps 1;

```

```

QY 1 MCKFWMAWGLCLVPLSLAQIDNLTCPAGVHYEKNGRYSISTEADLCKAFNSTL 60
DB 1 MCKFWMAWGLCLVPLSLAQIDNLTCPAGVHYEKNGRYSISTEADLCKAFNSTL 60
QY 61 PTMAQWEKALSIGFETCRVGFEGHVVIRIHPNSICANNVTGYILTSNTSOYDYTCFN 120
DB 61 PTMAQWEKALSIGFETCRVGFEGHVVIRIHPNSICANNVTGYILTSNTSOYDYTCFN 120
QY 121 ASAPPEEDCTSVTDLNPAEDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLNPAEDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
QY 181 GSSSERSSTGGIIFTFSTVHPIPEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240
DB 181 GSSSERSSTGGIIFTFSTVHPIPEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 240
QY 241 GLVEDLDRIGPLSMNTQOQNSQSFSTSHGLEEDKHPTTSLTSSNNDVTVGRRDPN 300
DB 241 GLVEDLDRIGPLSMNTQOQNSQSFSTSHGLEEDKHPTTSLTSSNNDVTVGRRDPN 300
QY 224 ----- 223
DB 224 ----- 223
QY 301 SEGSTLLEGYSHYPTKESRTFIPVTSKATGFGVAVTVGDSNVNRSLSGDQPTF 360
DB 224 -----DQDTF 228
QY 361 HPSGSGHTTHGSESDSHSGSQGGANTTSGPIRTPOIPEWLIILASLLALILAVCI 420
DB 229 HPSGSGHTTHGSESDSHSGSQGGANTTSGPIRTPOIPEWLIILASLLALILAVCI 288
QY 421 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASQSEMVHLVKNKSESETPDQMTAD 480
DB 289 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASQSEMVHLVKNKSESETPDQMTAD 348
QY 481 ETNQLQVDMKIGV 493
DB 349 ETNQLQVDMKIGV 361

```

RESULT 4
5504194-2

```

; Patent No. 5504194
; APPLICANT: ST. JOHN, THOMAS P.; GALLATIN, W. MICHAEL; IDZERDA,
; REBEAN
; TITLE OF INVENTION: LYMPHOCYTE ADHESION RECEPTOR FOR HIGH
; ENOHELIUM, CD44
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,624
; FILING DATE: 15-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 628,646
; FILING DATE: 12-DEC-1990
; APPLICATION NUMBER: 325,224
; FILING DATE: 17-MAR-1989
; SEQ. ID NO. 2:
; LENGTH: 362
; 5504194-2

```

```

Query Match 65.1%; Score 1698.5; DB 6; Length 362;
Best Local Similarity 68.2%; Pred. No. 1.3e-144;
Matches 337; Conservative 7; Mismatches 17; Indels 133; Gaps 2;

```

```

QY 1 MCKFWMAWGLCLVPLSLAQIDNLTCPAGVHYEKNGRYSISTEADLCKAFNSTL 60
DB 1 MCKFWMAWGLCLVPLSLAQIDNLTCPAGVHYEKNGRYSISTEADLCKAFNSTL 60
QY 61 PTMAQWEKALSIGFETCRVGFEGHVVIRIHPNSICANNVTGYILTSNTSOYDYTCFN 120
DB 61 PTMAQWEKALSIGFETCRVGFEGHVVIRIHPNSICANNVTGYILTSNTSOYDYTCFN 120
QY 121 ASAPPEEDCTSVTDLNPAEDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLNPAEDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
QY 181 GSSSERSSTGGIIFTFSTVHPIPEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 229
DB 181 GSSSERSSTGGIIFTFSTVHPIPEDSPWITDSTRIPATNMDSSHSTTLQPTANPT 224
QY 240 GLVEDLDRIGPLSMNTQOQNSQSFSTSHGLEEDKHPTTSLTSSNNDVTVGRRDPN 299
DB 225 ----- 224
QY 300 HSBGSTTLLEGYSHYPTKESRTFIPVTSKATGFGVAVTVGDSNVNRSLSGDQPT 359
DB 225 -----DQGA 228
QY 360 HPSGSGHTTHGSESDSHSGSQGGANTTSGPIRTPOIPEWLIILASLLALILAVCI 419
DB 229 HPSGSGHTTHGSESDSHSGSQGGANTTSGPIRTPOIPEWLIILASLLALILAVCI 288
QY 420 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASQSEMVHLVKNKSESETPDQMTAD 479
DB 289 VNSRRRCQKKLVINSNGAVEDRKPSGLNGEASQSEMVHLVKNKSESETPDQMTAD 348
QY 480 ETNQLQVDMKIGV 493
DB 349 ETNQLQVDMKIGV 362

```

RESULT 5
US-07-946-497-2
Sequence 2, Application US/07946497
Patent No. 5506119
GENERAL INFORMATION:
APPLICANT: HERBLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHER, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SRO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-946-497-2

Query Match 54.6%; Score 1425; DB 1; Length 503;

Best Local Similarity 58.7%; Pred. No. 9.6e-120; Indels 78; Gaps 16;
Matches 315; Conservative 36; Mismatches 108;

QY 1 MDKFWMAAANG-LCLVPLSLA--QIDNITCRFAGVHVHVKNGRYSISRTEADLCAFN 57
DB 1 MDKFWMAAANG-LCLVPLSLA--QIDNITCRFAGVHVHVKNGRYSISRTEADLCAFN 60
QY 58 STLPMAQMEKALSIGFTCYGFIIEGHVVIPIRHPSICAAANTGYI-LTSNTQYDT 116
DB 61 TLPPTMAQMEKALSIGFTCYGFIIEGHVVIPIRHPSICAAANTGYI-LTSNTSHYDT 120
QY 117 YCFNASAPREDCTSVTDLPNAPFGPITITIVNDRGTRYVOKGEYRINPEDIYPSNPTD 176
DB 121 YCFNASAPREDCTSVTDLPNAPFGPITITIVNDRGTRYVOKGEYRINPEDIYPSNPTD 180
QY 177 DVSSGSSSERSTSGGIYFT-FSTVHPIDED-----SPWITDST---D 217
DB 181 DVSSGSSSERSTSGGIYFT-FSTVHPIDED-----SPWITDST---D 217
QY 218 RIPATNMDSSTLTLOPTANPTGLVEDLR--TGPLSMTTQOQNSQSFSTSHGLEEDK 275
DB 240 RTQWNPPIHNSNEVLQTTTR-----MTDIDNSTSAGENWTOEPQPPFNHXYQDEET 294
QY 276 DHPFTSTLTSSNRDVTG-----GRDPMHSEGSTLLLEGY--SHYPTKE 320
DB 295 PHATSTWADPNSTTEEAATQKEMFENWQGNKPPSPSEDS--HVLEGTTASAHNNHPQ 353
QY 321 SRFTPIVTSAKTSGFVTAIVVGDNSGNVRSLSGDDDTF---HPGSGSHTHGSESDQ 376
DB 354 RMT-----TOSQEDVSWT-----DFPDPIISHPMQGHOT---ESKG 386
QY 377 HSHSQSGEGANTTSGPIRTPOIPFWIIILASLALAILAVCIAVNSRRRCQKXKLVIN 436
DB 387 HSSGADSGVTTTGGPARPQIPFWIIILASLALAILAVCIAVNSRRRCQKXKLVIN 446
QY 437 SGNQAVEDRKPSGANGASKQEMVHLVKNESSETPQPMFADETRLQAVDKIGV 493
DB 447 SGNQAVEDRKPSGANGASKQEMVHLVKNESSETPQPMFADETRLQAVDKIGV 503

RESULT 6

US-08-483-322-2
Sequence 2, Application US/08483322
Patent No. 5760178

GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHER, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD4 SURFACE PROTEINS, DNA
SEQUENCES CODING THERE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,322
FILING DATE: 07-0UN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,497
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

INFORMATION FOR SRO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-322-2

Query Match 54.6%; Score 1425; DB 1; Length 503;
Best Local Similarity 58.7%; Pred. No. 9.6e-120; Indels 78; Gaps 16;
Matches 315; Conservative 36; Mismatches 108;

QY 1 MDKFWMAAANG-LCLVPLSLA--QIDNITCRFAGVHVHVKNGRYSISRTEADLCAFN 57
DB 1 MDKFWMAAANG-LCLVPLSLA--QIDNITCRFAGVHVHVKNGRYSISRTEADLCAFN 60
QY 58 STLPMAQMEKALSIGFTCYGFIIEGHVVIPIRHPSICAAANTGYI-LTSNTQYDT 116
DB 61 TLPPTMAQMEKALSIGFTCYGFIIEGHVVIPIRHPSICAAANTGYI-LTSNTSHYDT 120
QY 117 YCFNASAPREDCTSVTDLPNAPFGPITITIVNDRGTRYVOKGEYRINPEDIYPSNPTD 176
DB 121 YCFNASAPREDCTSVTDLPNAPFGPITITIVNDRGTRYVOKGEYRINPEDIYPSNPTD 180
QY 177 DVSSGSSSERSTSGGIYFT-FSTVHPIDED-----SPWITDST---D 217
DB 181 DVSSGSSSERSTSGGIYFT-FSTVHPIDED-----SPWITDST---D 217
QY 218 RIPATNMDSSTLTLOPTANPTGLVEDLR--TGPLSMTTQOQNSQSFSTSHGLEEDK 275
DB 240 RTQWNPPIHNSNEVLQTTTR-----MTDIDNSTSAGENWTOEPQPPFNHXYQDEET 294
QY 276 DHPFTSTLTSSNRDVTG-----GRDPMHSEGSTLLLEGY--SHYPTKE 320
DB 295 PHATSTWADPNSTTEEAATQKEMFENWQGNKPPSPSEDS--HVLEGTTASAHNNHPQ 353

Db 295 PHATSTTWADPNSTTEEAATQKEKPFENEMOGKNPPTSEDS-HYEGTTASAHNNHPSQ 353
Qy 321 SRTFIPVTSKATGSPGVTAVTVGDSNNVNRSLSGDQTF---HPGGSHTTHGSESDG 376
Db 354 RMT-----TQSGEDVSWT-----DFPDPISHMGQGHOT---ESKG 386
Qy 377 HSHSQGEGANTTSGPIRTPQIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 436
Db 387 HSSGNODSGVTTTSGPARPQIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 446
Qy 437 SGNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDQPMTADETRLQNVDMKIGV 493
Db 447 SGNGVEDRKPSGLNGEASKSQEMVHLVNXKEPTETPDQPMTADETRLQNVDMKIGV 503

RESULT 7
US-08-478-882-2
Sequence 2, Application US/08478882
Patent No. 5885575
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: FONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,882
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-882-2

Query Match 54.6%, Score 1425, DB 2; Length 503;
Best Local Similarity 58.7%, Pred. No. 9.6e-120;
Matches 315; Conservative 36; Mismatches 108; Indels 78; Gaps 16;

Qy 1 MDKFWTHAANG-LCLVPLSLA--QIDNITRCFAGVPHVENGRYSISRTAADLCAFN 57
Db 1 MDKFWTHAANG-LCLVPLSLA--QIDNITRCFAGVPHVENGRYSISRTAADLCAFN 60
Qy 58 STLTPTMAOMKALIGFTCYGRIEGHVIVPRHNSICAAANTGYIT-LTENTGYDT 116

Db 61 TLTPTMAOMKALIKGEFTRCYGRIEGHVIVPRHNSICAAANTGYITLANTSHYDT 120
Qy 117 YCFNAPSAPREDCTSVTLDPNAFDGPTITITIVNRDGRVYQKGEYRINPEDIVPSNPTD 176
Db 121 YCFNAPSAPREDCTSVTLDPNPFDPVITITIVNRDGRVYQKGEYRINPEDIVPSNPTD 180
Qy 177 DVSSGSSSEASISGCGIIFPT-ESTVHPIDED-----SPWITST---D 217
Db 181 DVSSGSSSEASISGCGIIFPT-ESTVHPIDED-----SPWITST---D 217
Qy 218 RIPATNDSHSTLQPTANPTGLVVDLDR--TGPLSMTQOSNSQSFSTSHGLGEEDK 275
Db 240 RTQNNPHNSNEVLLQTTT-----MTDIDRNSASANGEMWTGEPQPFNNHXYQDEET 294
Qy 276 DHPTSTLTSSNRDVTG-----GRDDPNHSGSTTLLEGYT--SHYPTKE 320
Db 295 PHATSTTWADPNSTTEEAATQKEKPFENEMOGKNPPTSEDS-HYEGTTASAHNNHPSQ 353
Qy 321 SRTFIPVTSKATGSPGVTAVTVGDSNNVNRSLSGDQTF---HPGGSHTTHGSESDG 376
Db 354 RMT-----TQSGEDVSWT-----DFPDPISHMGQGHOT---ESKG 386
Qy 377 HSHSQGEGANTTSGPIRTPQIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 436
Db 387 HSSGNODSGVTTTSGPARPQIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 446
Qy 437 SGNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDQPMTADETRLQNVDMKIGV 493
Db 447 SGNGVEDRKPSGLNGEASKSQEMVHLVNXKEPTETPDQPMTADETRLQNVDMKIGV 503

RESULT 8
US-07-946-497-7
Sequence 7, Application US/07946497
Patent No. 5506119
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: FONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: MCD44
US-07-946-497-7

Query Match 52.0%; Score 1357; DB 1; Length 363;
Best Local Similarity 57.9%; Pred. No. 7.6e-114;
Matches 288; Conservative 28; Mismatches 43; Indels 138; Gaps 9;

1 MDKFWMAWGLCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFNS 58
1 MDKFWMAWGLCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFNS 60
59 TLPTMAQMEKALSGFETCRGFEHGVVIRIHPNSICANNVGYIL-TSNTSOYDTY 117
61 TLPTMDQKALSKGFETCRGFEHGVVIRIHPNACANHGVIYILVTSNTSHDTY 120
118 CFNASAPPEEDCTSVTDLFNAFDGPITITIVNRDGTTRYVKGERTNPEDIYPSNPTDD 177
121 CFNASAPPEEDCTSVTDLFNSFDGPVITITIVNRDGTTRYVKGERTNPEDIYPSNPTDD 180
178 VSSGSSSRSTSGYIYTFSTVHPFIPDEDSFWITDSTRIPATNMDSHSTLQPTAN 237
181 VSSGSTIEK-STPEGYILHTY---LPTB-----QPTG- 208
238 PNTGLVEDLDRTGLSMTQOQNSQSFSTSHGLEEDKHPTTSTLSSNNDVTGARD 297
209 -----DQDSSFIRSTLATRDSDS----- 227
298 PNHSEGSTTLLEGYTSHPHTKESRTFIPVTSAGTSFGVTAATVGDNSNVNRLSGDQ 357
228 -----SKDSR----- 232
358 DTFHPSGSGHT-THGSESDGSHSGOEGANTTSGPIRTPOIPEWIIILASLALALILA 416
233 -----GSSRTVTHGSELAGHSANQDSGVTTTSGPMRPOIPEWIIILASLALALILA 286
417 VCIANVSRRCGQKKLVINGNGAVEDRKPSGLNGASYSQEMVHLVNXESSPTDQFM 476
287 VCIANVSRRCGQKKLVINGNGAVEDRKPSGLNGASYSQEMVHLVNXESSPTDQFM 346
477 TADETRNLQNVDMKIGV 493
347 TADETRNLQNVDMKIGV 363

RESULT 9

US-08-483-322-7
Sequence 7, Application US/08483322
Patent No. 5760178
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,322
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,497
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5359
TELEX: 904136

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: MCD44
US-08-483-322-7

Query Match 52.0%; Score 1357; DB 1; Length 363;
Best Local Similarity 57.9%; Pred. No. 7.6e-114;
Matches 288; Conservative 28; Mismatches 43; Indels 138; Gaps 9;

1 MDKFWMAWGLCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFNS 58
1 MDKFWMAWGLCLVPLSLA--QIDNITCRFAGVHVEKNGRYSISRTAADLCKAFNS 60
59 TLPTMAQMEKALSGFETCRGFEHGVVIRIHPNSICANNVGYIL-TSNTSOYDTY 117
61 TLPTMDQKALSKGFETCRGFEHGVVIRIHPNACANHGVIYILVTSNTSHDTY 120
118 CFNASAPPEEDCTSVTDLFNAFDGPITITIVNRDGTTRYVKGERTNPEDIYPSNPTDD 177
121 CFNASAPPEEDCTSVTDLFNSFDGPVITITIVNRDGTTRYVKGERTNPEDIYPSNPTDD 180
178 VSSGSSSRSTSGYIYTFSTVHPFIPDEDSFWITDSTRIPATNMDSHSTLQPTAN 237
181 VSSGSTIEK-STPEGYILHTY---LPTB-----QPTG- 208
238 PNTGLVEDLDRTGLSMTQOQNSQSFSTSHGLEEDKHPTTSTLSSNNDVTGARD 297
209 -----DQDSSFIRSTLATRDSDS----- 227
298 PNHSEGSTTLLEGYTSHPHTKESRTFIPVTSAGTSFGVTAATVGDNSNVNRLSGDQ 357
228 -----SKDSR----- 232
358 DTFHPSGSGHT-THGSESDGSHSGOEGANTTSGPIRTPOIPEWIIILASLALALILA 416
233 -----GSSRTVTHGSELAGHSANQDSGVTTTSGPMRPOIPEWIIILASLALALILA 286
417 VCIANVSRRCGQKKLVINGNGAVEDRKPSGLNGASYSQEMVHLVNXESSPTDQFM 476
287 VCIANVSRRCGQKKLVINGNGAVEDRKPSGLNGASYSQEMVHLVNXESSPTDQFM 346
477 TADETRNLQNVDMKIGV 493
347 TADETRNLQNVDMKIGV 363

RESULT 10

US-08-478-882-7
Sequence 7, Application US/08478882
Patent No. 5885575
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

Db 209 -----GDR 211
QY 297 DPNHSGSTLLBGLYSHYPTKESRTFFIPVTSAGTSFGVAVTVVGDNSNVNRSLSGD 356
Db 212 DDAFFIGSTL----- 221
QY 357 QDFHPSGSHHTHGSSEDSHSGHSGOGANTTSGPIRTPQIPENTIIILASLALALILA 416
Db 222 -----ATGHSNGNDSSGVTTTSGPARRQIPENIIILASLALALILA 264
QY 417 VCIANRRRCGCKKLVINSNGAVEDRKPSGLNGBASCSQMVLVNKESSETPDQM 476
Db 265 VCIANRRRCGCKKLVINSNGVTRKPSGLNGBASCSQMVLVNKESSETPDQM 324
QY 477 TADETRLQNVDMKI 491
Db 325 TADETRLQNVDMKI 339
RESULT 12
US-07-946-497-4
; Sequence 4, Application US/07946497
; Patent No. 5506119
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-946-497-4
Query Match 27.3%; Score 711; DB 1; Length 354;
Best Local Similarity 70.0%; Pred. No. 1,1e-55;
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;
QY 161 YRTNEDIVPSNPTDDVSSGSSSRSTSGGYFYRTSTVHPI-----PDDESPWITD 214
Db 161 YRTNEDIVPSNPTDDVSSGSSSRSTSGGYFYRTSTVHPI-----PDDESPWITD 201
QY 215 STDRIIP-----ATNMDSHSTTLQPTANPNTGLVEDLDRGTGLSMVTOOSNSG 263

Db 202 FPNPISHPMGRGHQAGRBMWDSHSTTLQPTANPNTGLVEDLDRGTGLSMVTOOSNSG 261
QY 264 FSTHGELEEDKHPTTSTLTSNRNDVYTGRRDPNHSBSTTLLEGYSHYPTKESRT 323
Db 262 FSTHGELEEDKHPTTSTLTSNRNDVYTGRRDPNHSBSTTLLEGYSHYPTKESRT 321
QY 324 FIPVTSAGTSFGVAVTVVGDNSNVNRSLSGD 356
Db 322 FIPVTSAGTSFGVAVTVVGDNSNVNRSLSGD 354
RESULT 13
US-08-483-322-4
; Sequence 4, Application US/08483322
; Patent No. 5760178
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,322
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,497
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-322-4
Query Match 27.3%; Score 711; DB 1; Length 354;
Best Local Similarity 70.0%; Pred. No. 1,1e-55;
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;
QY 161 YRTNEDIVPSNPTDDVSSGSSSRSTSGGYFYRTSTVHPI-----PDDESPWITD 214
Db 161 YRTNEDIVPSNPTDDVSSGSSSRSTSGGYFYRTSTVHPI-----PDDESPWITD 201
QY 215 STDRIIP-----ATNMDSHSTTLQPTANPNTGLVEDLDRGTGLSMVTOOSNSG 263
Db 202 FPNPISHPMGRGHQAGRBMWDSHSTTLQPTANPNTGLVEDLDRGTGLSMVTOOSNSG 261
QY 264 FSTHGELEEDKHPTTSTLTSNRNDVYTGRRDPNHSBSTTLLEGYSHYPTKESRT 323

Db 262 FSTSHGLEEDKCHPTTSTLTSSNRNDVVGRRDPNHSSEGTLLLEGYTSHPHYKESRT 321
QY 324 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 356
Db 322 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 354

RESULT 14
US-08-478-882-4
Sequence 4, Application US/08478882
Patent No. 5885575
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHER, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD4 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,882
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-882-4

Query Match 27.3%; Score 711; DB 2; Length 354;
Best Local Similarity 70.0%; Pred. No. 1.1e-55;
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

QY 161 YRTNPEDLYPSNPDDVDVSGSSERSSTGCGYFFYFSTVHPI-----PDEDSPWITD 214
Db 161 YRQTPRE-----DSHSTGTAAASAHTS-----HPQGRTPSPEDSSW-TD 201

QY 215 STRIP-----ATNNDSSHSTTLOPTANPNTGLVEDLDRGTGLSMTQOOSNQS 263
Db 202 FFPNISHPMGRGHQAGRMDSDSHSTTLOPTANPNTGLVEDLDRGTGLSMTQOOSNQS 261

QY 264 FSTSHGLEEDKCHPTTSTLTSSNRNDVVGRRDPNHSSEGTLLLEGYTSHPHYKESRT 323
Db 262 FSTSHGLEEDKCHPTTSTLTSSNRNDVVGRRDPNHSSEGTLLLEGYTSHPHYKESRT 321

QY 324 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 356

Db 322 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 354

RESULT 15
US-08-359-850-2
Sequence 2, Application US/08359850
Patent No. 5951982
GENERAL INFORMATION:
APPLICANT: Ziller, Margot
APPLICANT: Herlich, Peter
APPLICANT: Ponta, Helmut
TITLE OF INVENTION: Use of Antibody-Containing Preparations
TITLE OF INVENTION: for Immunosuppression
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox
STREET: 1225 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/359,850
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/963,323
FILING DATE:
APPLICATION NUMBER: DE P 41 34 982.2
FILING DATE: 23-OCT-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 833-7553
TELEFAX: (202) 833-8716
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 338 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-359-850-2

Query Match 26.8%; Score 699; DB 2; Length 338;
Best Local Similarity 69.7%; Pred. No. 1.2e-54;
Matches 147; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

QY 161 YRTNPEDLYPSNPDDVDVSGSSERSSTGCGYFFYFSTVHPI-----PDEDSPWITD 214
Db 147 YRQTPRE-----DSHSTGTAAASAHTS-----HPQGRTPSPEDSSW-TD 187

QY 215 STRIP-----ATNNDSSHSTTLOPTANPNTGLVEDLDRGTGLSMTQOOSNQS 263
Db 188 FFPNISHPMGRGHQAGRMDSDSHSTTLOPTANPNTGLVEDLDRGTGLSMTQOOSNQS 247

QY 264 FSTSHGLEEDKCHPTTSTLTSSNRNDVVGRRDPNHSSEGTLLLEGYTSHPHYKESRT 323
Db 248 FSTSHGLEEDKCHPTTSTLTSSNRNDVVGRRDPNHSSEGTLLLEGYTSHPHYKESRT 307

QY 324 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 354
Db 308 FIPVTSKATGSPGVTAIVTVDGNSNVNRSLSGD 338

Search completed: March 8, 2004, 06:10:38
Job time : 26 secs